WARREN H. FAY

Temporal Sequence in the Perception of Speech

MOUTON & CO.

TEMPORAL SEQUENCE
IN THE PERCEPTION OF SPEECH

JANUA LINGUARUM

STUDIA MEMORIAE
NICOLAI VAN WIJK DEDICATA

edenda curat

CORNELIS H. VAN SCHOONEVELD
STANFORD UNIVERSITY

SERIES MINOR
NR. XLV

1966
MOUTON & CO.
LONDON · THE HAGUE · PARIS

TEMPORAL SEQUENCE IN THE PERCEPTION OF SPEECH

by

WARREN H. FAY

UNIVERSITY OF OREGON
MEDICAL SCHOOL

1966

MOUTON & CO.

LONDON · THE HAGUE · PARIS

Printed in The Netherlands by Mouton & Co., Printers, The Hague.

To Carol Fay
my wife and collaborator

PREFACE

This book brings together from a wide variety of sources information relative to a rarely investigated but highly pertinent dimension of oral language, temporal sequence. Broadbent (10, p. 47) has said, "(Humans) can use language because they can deal with sequences ... and speech is the most obvious case of stimuli being dealt with in sequences." In addition to a review and discussion of an extensive cross-disciplinary literature dealing with theoretical and experimental contributions (Part I), an investigation of temporal resolution of phonemic elements and pure tones is presented for its experimental methodology and potentially useful data (Part II). Finally, in Part III, conclusions based upon the data of the previous sections are briefly summarized together with some theoretical implications.

If, initially, I was not cognizant of the complexities involved in the study of human auditory perception and the elusive primary variable, time, I am now humbled by their magnitude. Perhaps the late Norbert Weiner (95) stated the problem best in these words:

> When we are dealing with man or an animal as the receptor we are in a very mixed situation. We are dealing with a receptor of whose general character we may know a good deal, of whose particular hitch-up we know very little. The result is that the sort of thing that becomes a clear cut quantitative statement in machine communication becomes a qualitative statement in linguistics ... Our philologists and engineers both have the problem of studying the properties of a little black box with terminals where they don't know what is inside. This is a much less important problem for the engineer than for the philologist. The engineer, after all, can rip his box open. The philologist can't.

Nevertheless, the philologist, the physicist, the physiologist, the psychologist, the physician, the speech pathologist and many other representatives of scientific endeavor are confronting that black box each from the framework of his particular discipline. Some workers have inferred the operations through controlled input whereas others have explored only the output; some have ablated or rewired the black box whereas others have begun their studies with damaged apparatus; and some have simply philosophized about its mysterious performance. It is hoped that the cross-disciplinary coverage of the problem will be useful to the professional who finds himself a layman when removed from his specialty, and, more particularly, to the professional whose scope excludes the potential of other scientific disciplines. My approach is often that of a curious layman and as such, I must apologize to those whose pertinent contributions I may have overlooked or slighted or whose discipline I have inadvertently failed to explore.

I wish to acknowledge my deepest appreciation to my mentor, Dr. T. D. Hanley, whose guidance and assistance made this book a reality. I am also indebted to Dr. M. D. Steer, director of the Purdue University Department of Audiology and Speech Sciences where the experimental portions of this study were completed. His support and personal encouragement were freely given and gratefully received.

I wish also to thank Dr. J. T. Graham for his assistance in the area of audition and Dr. C. D. Smock for his instruction in subjects of time and human perception. Without the design and data-handling advice of Dr. K. W. Burk and the technical assistance of Merle Pheil, obstacles of this project would have been enormous if not defeating; with their help these problems were challenges. Finally, I acknowledge the welcome financial support of the Purdue Research Foundation.

W. H. F.

TABLE OF CONTENTS

PART III

APPENDICES

PART I
AN OVERVIEW

THEORETICAL CONSIDERATIONS

THE ACOUSTIC THEORISTS

Through the detailed study of the sound wave and its psychoacoustic properties, an extensive area of experimental and theoretical endeavor has developed. This section deals with the theoretical backgrounds for much of the time-oriented auditory research recently undertaken.

With his "Space-Time Theory of Hearing" Fletcher (27, p. 276) describes the translation of the sound wave into a form suitable for transmission by the nervous system. According to his theory the pitch of a tone is determined by the position of maximum stimulation on the basilar membrane of the cochlea and also by the time pattern sent to the brain. (Fletcher postulates that the former is more important for the higher tones and the latter for the low tones.) Loudness is determined by the number of impulses per second reaching the brain and the extent of the stimulated area. The time pattern in the air is converted into a space pattern on the basilar membrane. The nerve endings are excited in such a way that this space pattern is transferred from the left ear, for example, to the brain, producing two space patterns which are almost alike, one on the left side and one on the right side. Similarly, a sound in the right ear produces two patterns, one on the left and the other on the right. Thus, when listening to a sound with both ears, there are four space patterns produced in the brain. Enough of the time pattern in the air is sent to each of these stimulated patches to make the time of maximum stimulation in each patch detectable. "It is a recognition of the changes in these patterns", Fletcher concludes, "that accounts for all the phenomena of audition."

Temporal aspects both of the incoming signal and of the neural transmission of the auditory pattern are considered by Wever (97, p. 440) in his "Volley Theory of Hearing". He maintains that there are almost unlimited ways in which time can enter into the pattern of action of a sound. Among the temporal phenomena that he analyzes are variations of amplitude, frequency, and phase. Wever theorizes that every neural element relays two sorts of information, the position along the membrane that it serves and the periodicity of its excitation. Although various conditions along the auditory tract combine in broadening and blurring the represented pattern, enough remains of the form of the pattern to join with the frequency representation in auditory perception.

Fletcher's and Wever's conceptions of partial preservation of the temporal factors as the signal ascends the neural pathways is shared, in general, by Licklider (55). The latter theorist, however, suggests that in speech perception there is an early stage in which the stream of speech is segmented into elements, and the elements are recognized or identified. He considers this process not logically necessary, but says "... there is so much evidence that speech is basically a sequence of discrete elements that it seems reasonable to limit consideration to mechanisms that break the stream down into elements and identify each element as a member of one or another of a finite number of sets." Because time cannot be preserved through many synapses, this stage of the auditory process is designed to take advantage of the time detail, according to Licklider. He concedes, however, the possibility of a broad-band pathway by-passing the lower stages and carrying as much temporal detail as the nervous system can retain all the way to the cerebral cortex. The Licklider theory goes on to suggest that recognition of phonemes might involve matched filtering whereas the comprehension of ideas might involve correlation.

Contrasting the *events* of audition with the *objects* of visual perception, Hirsh (39) held in 1952 that time is the dimension within which patterns are articulated for hearing in a manner analogous to the way in which space is the patterning dimension for vision. In a subsequent paper, 1959, Hirsh (41) concluded that temporal resolution of order requires additional mechanisms than those associated with the peripheral auditory system.

Central involvement in the analysis of the signal is also emphasized by Davis (19, p. 139). He describes the auditory system as dependent upon an elaborate set of codes for transmitting the information concerning frequency, intensity, and time differences that is contained in the incoming acoustic signals. He further theorizes that the multiplicity of codes implies an equal number of central receiving or decoding mechanisms to make central integration possible.

Huggins (44) hypothesized a system-function analysis of speech sounds. He suggested that our own consciousness is built around a reference frame which moves with time such that the present is always immediately in front of us and our awareness of events as they slide by into the past is continually changing. His theory postulated that in some manner the auditory mechanism is able to perform a generalized frequency analysis upon the system function associated with particular speech sounds rather than upon the sound wave itself. In addition, excitation pulses are analyzed for their temporal characteristics and periodicities – perhaps by a sort of autocorrelation representation.

In a subsequent publication, Huggins (45) suggested a phase principle which he said permits effective use of the available neural elements and their mechanisms of synaptic facilitation and inhibition. Huggins stated the principle as follows:

> Because each time interval is independent of the amplitudes of the waves, its accurate measurement is easily possible over very wide dynamic ranges of signal intensity provided only that zero crossing can be detected accurately. Regardless of the type of analysis occurring in the ear, the results must ultimately be delivered to the central nervous system in the form of "all or none" impulses. Such neural messages are beautifully able to convey temporal data, such as the instants of zero crossings; they are poorly suited for conveying amplitude and intensities.

Licklider and Miller (57) and Hirsh (41) also discussed the temporal aspects of phase, but the latter author emphasized that temporal relations as expressed in terms of frequency and phase tend to "stay put" for as long as the sound is produced. Hirsh pointed out that temporal differences involved in a single acoustic event are extremely small.

In summary, the acoustic theorists are primarily concerned with the temporal phenomena embedded within the auditory signal and the fidelity with which the parameter is preserved in the ascending neural pathways. All the theorists cited credit both the peripheral and central processes in the temporal coding and preserving operations, but most of the writers are limited to discussing variables utilized in the central processes rather than the processes themselves. There is a strong suggestion in the works of some theorists that speech perception may be a separate and somewhat distinct function from that of acoustically less complex auditory input. Time, nevertheless, is extensively involved in the many phases of the total auditory process, be it in the phase and frequency variations, duration of input signal, periodicity of neural excitation, segmentation of the sound stream into elements, or the time of maximum stimulation of cerebral space patterns.

THE INFORMATION THEORISTS

A sizeable number of authorities have applied the principles of information theory to language communication. Miller (64), Attneave (4), Shannon and Weaver (80), Weiner (94), Fano (23), Stroud (85), Frick (28), and Cherry (16) are among those who have considered speech in this manner. Briefly, the information theory view holds that the elements of the input signal are chained sequentially under the influence of conditional probabilities. The process of reception takes into account both new evidence and the *a priori* information contained in the statistics of the already received signal. As information continues to come in, the listener narrows the possibilities until there is a sufficiently strong concentration of likelihood about one phoneme or sequence of phonemes to warrant the decision that is the correct one. It is then stored, and helps to condition the calculations for the next condition (56).

The theory of information is based, according to Fano (23), on two fundamental premises, one quantistic and the other statistical. The quantistic concept assumes that any process conveying information consists inherently of a sequence of indications of one choice out of a number of possible choices. Now the question arises, how can the speechwave be at once both a continuous acoustical stream bound to the time continuum and a series of sequentially-linked sound units, informational units, and linguistic units? The answer to this question depends largely upon the purpose and frame of reference of the responder.

By way of definition Shannon and Weaver (80, p. 6) describe a discrete system as one in which both message and signal are a sequence of discrete symbols; a continuous system is one in which both message and signal are treated as continuous functions.

Joos (47), a linguist, lends perspective to the issue by pointing out that physicists describe speech with continuous mathematics, such as Fourier analysis or the autocorrelation function. Linguists, on the other hand, describe language using a discontinuous or discrete mathematics called "linguistics". Joos emphasizes that it is the duty of the linguist to describe precisely, not to explain: "And we feel that our descriptive statements fit actual speech behavior, but we have no right to claim that they are 'correct' in the sense that they fit the neural events in the brains of the speaker and listener." Joos defines the mathematics of linguistics as follows:

> Ordinary mathematical techniques fall mostly into two classes, the continuous (e.g. the infinitesimal calculus) and the discrete or discontinuous (e.g. finite group theory). Now it will turn out that the mathematics called "linguistics" belongs to the second class. It does not even make any compromise with continuity, as does statistics, or infinite-group theory. Linguistics is a quantum mechanics in the most extreme sense. All continuity, all possibilities of infinitesimal gradation, are shoved outside of linguistics in one direction or the other. There are in fact two such directions in which we can and resolutely do expel continuity; semantics and phonetics.

Thus, it may be seen that the limits defining a continuous or segmented system are dependent in part upon the particular professional discipline of the investigator. Inasmuch as the area of communication is studied by a diversity of professions, it is reasonable that cross-disciplinary investigation may become confused if the limits imposed by a given discipline are not widely known and understood. In order to provide for ease of comparison and condensation of an extensive body of literature, a number of views on the issue of *continuous* and *discrete* in relation to perception of the flowing speech signal have been assembled in Table I.

Speech perception theorists such as Licklider (55), Huggins (44), and Miller (64) attempt to explain the process rather than to describe it. These writers suggest that the speech flow is segmented into units by the human system. Miller (64) observes, "There are few clues in the physical process to indicate how this continuous stream of sound is to be segmented, yet every speaker and listener deals with the stream as though it consisted of a sequence of isolated elements put together like beads on a string."

Fant (24) has summarized schematically speech wave segmentation from several points of view. This conception of segmentation may be seen in Figure 1. Model a) is a sequence of ideal non-overlapping phonemes whereas b) is a sequence of minimal sound segments, the boundaries of which are defined by relative distinct changes in the speech wave structure. It will be noted that the number of successive sound segments within an utterance is greater than the number of phonemes. Model c) shows that one or more of the sound features characterizing a sound segment may extend over several boundaries as, for example, the influence exerted by a consonant on a following vowel. The d) model shows a continuously varying importance function for each phoneme describing the extent of its dependency on particular events within the speech wave. This model shows overlapping curves without sharp boundaries. Fant maintains that although the models may represent quite different views of the nature of speech, they are not contradictory in any way.

TABLE I

Theoretical considerations of the segmentation process in the perception of continuous speech

Theorist	*Source*	*Segmentation Views*
Cherry	(16) p. 188	"Continuous" functions are the creation of mathematicians for describing *physical signals*; mathematicians deal with mental *constructs*, not with description of physical situations.
Fischer-Jørgensen	(26)	It is impossible to cut sounds transversely to the time axis; there is overlapping of the sound features belonging to different phonemes.
Lisker	(59)	Segmentation need not account for every feature that contributes to the identification of a linguistic element.
Broadbent	(10) p. 279	The waveform of speech forms a sequence of events to which discrete responses are made. There may be several successive changes take place before a decision is reached.
Peterson	(73)	Segmentation directs major attention to unit boundaries and to locations when sectioning occurs; quantization directs attention to the central structure of the element, allowing vague boundaries.
Hockett	(43)	The signal is a sequence of moveable boundaries determined by the conditioning effect of the previous input.
Fairbanks	(22)	The elements of speech are not produced in stepwise fashion like the notes of a piano, but by continuous modulation as a function of time.
Fant	(24)	Speech can be divided into a sequence of sound segments the acoustic boundaries of which are definable either from specific articulatory events or from the corresponding time-selective changes in the spectral composition of the speech wave.

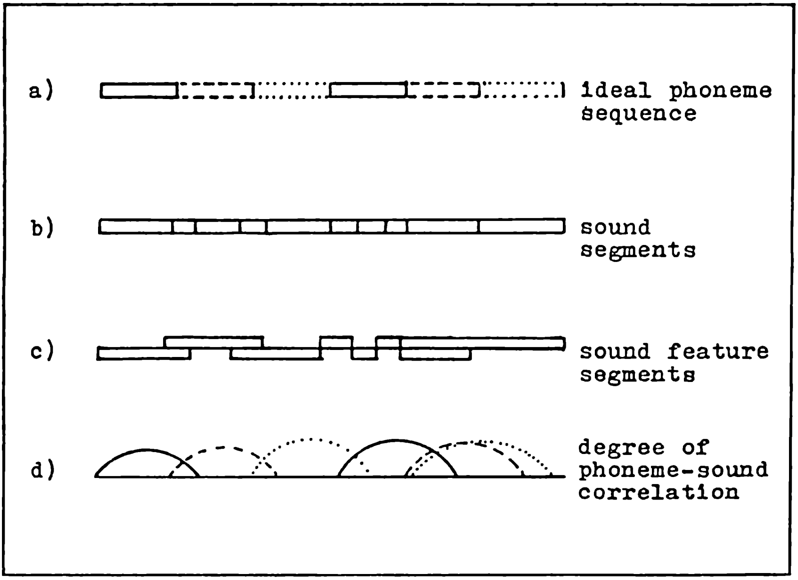


Figure 1. Schematic representation of sequential elements of speech; a) is the phonemic aspect, b) and c) represent acoustic aspects, and d) shows the degree of phoneme-sound correlation (After Fant [24]).

The overlap in the time domain according to d) does not invalidate the concept of the phonemes as discrete and successive in a). The representation in a) relates to the message aspect of speech communication whereas representations b) and c) pertain to the speech wave and d) more to the perception of speech.

Summarizing, the application to speech of information theory and its concept of sequential dependencies takes into account both the function of memory in language use and the *a priori* information contained in the time-flowing signal. For an event or events to be sequential in nature implies that some system must act upon a steady inflow (speech wave) segmenting or quantizing and perhaps patterning it into more or less discrete perceptual units or temporal packages suitable for analysis by the system. The theoretical assumptions of the manner by which segmentation occurs within the system are dependent to some extent on the perspective of the segmenting theorist. Information theory, therefore, is not so much concerned with the temporal aspects of the physical sound, per se, as it is with the units of information sequentially-linked to symbolically transmit a message by sound wave alterations. Aspects basic to the problem of understanding such a system include (a) the duration or size of the informational units, (b) the amount of overlap or the size of the inter-unit intervals, (c) the specific information cues within a given unit, and (d) the extent to which a given temporal unit is dependent upon the information preceding and succeeding its entrance into the system.

THE PHYSIOLOGICAL THEORISTS

Though all the theorists acknowledge the extensive role of the human system in the communication process, some more than others place considerable emphasis upon the human physiological timing of the sensory and motor operations involved. Lindsley (58), who believes that time is perhaps the most useful parameter in transforming and integrating behavioral and physiological data, has developed a "neuronic shutter" theory based upon his electroencephalographic research. He asserts that time and its derivatives seem to be fundamental properties or conditions. In Lindsley's theory the alpha activity cycle may be a means of coding or timing sensory impulses, "in order that our perceptual world and our reactions to it are not distorted or smeared by the more or less continuous influx of sensory stimuli". The proposed mechanism operates on the principle that the probability that incoming impulses will cause a neuron to fire will vary with the phase of the excitability (alpha) cycle. Impulses arriving at synapses when the trans-synaptic neuron is in the phase of increased excitability will be more likely to fire the trans-synaptic neuron. Those arriving during the phase of lowered excitability will be less likely to do so. And when the excitability cycles of a group of neurons are synchronized, then the flow of impulses through the group will be timed by the frequency and phase of the cycle.

An incomplete physiological theory that was explicitly formulated to account for temporal integration has been published by Lashley (50, chap. 4). Temporal integration apparent in language, according to this theory, is also evidenced in such motor functions as leg movements in insects, the song of birds, the control of trotting and pacing in a gaited horse, and in the movements of the carpenter sawing a board. Lashley also points out that reproductive memory and even visual scanning are reflected as systems of temporal sequence. To illustrate his concept of serial order in language, Lashley (50, p. 183) writes:

> Let us start the analysis of the process with the enunciation of the word. Pronunciation of the word "right" consists first of retraction and elevation of the tongue, expiration of air and activation of the vocal cords; second, depression of the tongue and jaw; third, elevation of the tongue to touch the dental ridge, stopping of vocalization, and forceful expiration of air with depression of the tongue and jaw. These movements have no intrinsic order of association. Pronunciation of the word "tire" involves the same motor elements in reverse order. Such movements occur in all permutations. The order must therefore be imposed upon the motor elements by some organization other than direct association connections between them. So, for individual movements in writing or typing the word, finger strokes occur in all sorts of combinations. No single letter invariably follows *g*, whether *gh*, *ga*, or *gu* is written depends upon a set for a larger unit of action, the word.

Lashley concludes that because the elements of the sentence are readied or partially activated before the order is imposed upon them in expression, a scanning mechanism must be operating to regulate their temporal sequence.

Considerable progress has been made in recent years in ablation and electrophysiological studies of temporal discrimination abilities of animals. Neff (71), who has been a prime investigator in this type of study, explicitly states his view that certain aspects of language behavior can be investigated in animals other than man. He presents evidence that neural systems that are essential for pattern perception and serial motor activity can be identified through experiments on lower animals, and neurophysiological events which occur during these kinds of behavior can be recorded and examined.

As for the cytoarchitectonic locale of the temporal ordering processes, Neff (71) asserts the necessary role played by the cerebral cortex whereas other authorities credit brainstem areas with the primary responsibility. The lemniscal system of the brain was found by Poggio and Mountcastle (75) to be "endowed with an exquisite capacity for temporal and spatial discriminations" in sub-human studies. Amassian and Waller (2) have concluded from their animal studies that even in some rather primitive parts of the mammalian brainstem some sensory information seems to be subjected in part to temporal coding. They concur with Eccles in the view that perception depends upon a probability response of many neurons in which *a range of temporal patterns occurs*, rather than upon a "specific spatiotemporal" pattern of excitation. Livingston, Haugen and Brookhart (60) lend additional support to the importance of lower brain processing areas in their description of the functional organization of the central nervous system. They stress not only the interaction functions of the brainstem reticular formation with various other parts of the brain, but also the role played by this brain area in the spatiotemporal organization and segregation of different sensory modalities. These authors observe, "Indeed, there is so much activity going on in the reticular formation all the time and so many ways in which the sensory input can be modified there that one wonders why the time-space patterns of sensory impulses are not hopelessly scrambled before they reach the cortex".

In summary, three theoretical approaches to an understanding of the function of time in speech perception have been presented: the acoustic, the mathematical or informational, and the primarily physiological. Although these approaches are quite different, each may contribute to the identification of the processes involved. For example, if one considers a word such as *education* (or *noitacude*, for that matter) to be literally flowing into the ear like a length of yarn, the temporal properties of the sound of the word would provide the listener with valuable information about the uniqueness or similarity of the elements composing it, their relations to one another in time, and the overall distinctive acoustic characteristic of the sound complex. But if the listener had no language storage for purposes of comparison, it would not matter if the total sound complex produced the configuration *education*, *noitacude*, or even *aoicudetn*. On the other hand, assuming previous acquaintance with the inflowing *education*, the analysis of the brain upon the elements might accept a sequence of *tion* succeeding an *a* on the basis of conditioned probabilities of a single unit in its complex sound environment. Conversely, the sequence of *detn* would be rejected by the system as improbable, without information, or sequentially confused. Finally, that combination of sound alterations composing *education* must have had to be emitted in its linguistic beginning by the human vocal and articulatory apparatus. This apparatus, it is logical to assume, functions as a motor complex of sequential behavior following a model dictated by a pre-imposed temporal pattern adhering, in part at least, to the motor capabilities of the speech organs involved. This temporal pattern could, theoretically, impose a temporal sequence upon an incoming message as a set of conditioned probabilities based upon logical or typical motor speech sequences. Thus it could be speculated that each theoretical approach might be dependent upon and interrelated with the other two approaches with time the independent variable of each system.

EXPERIMENTAL EVIDENCE

Before judgments of temporal sequence can be made the listener must be aware of the successiveness of items or events (74, p. 294). With these perceptual stages in mind, the experimental literature has been divided first into the studies dealing with succession of auditory stimuli and, subsequently, into the ordering or patterning of the events.

TEMPORAL SUCCESSION

According to James (46, p. 613), Exner in 1875 distinctly heard the doubleness of two successive clicks of a Savart's wheel and the two successive snaps of an electric spark when their interval was made as small as about 1/500 of a second, 2 msec.

Since Exner's time a number of experimenters have undertaken the task of determining the human resolving power of various combinations of the click stimuli. In 1949 Wallach, Newman, and Rosenzweig (92) reported a series of experiments in which loudspeaker locations were varied to obtain delays. They concluded that the interval over which fusion takes place for sounds in close succession varied with the kind of sound. A temporal interval of as much as 5 msec resulted in fusion of single clicks, whereas the fusion interval between two more complex sounds was judged to be perhaps as long as 40 msec. Other studies dealing with pairs of clicks as summarized by Guttman, van Bergeijk, and David (34) indicated that temporal resolution appreciably begins to fail below about 8-10 msec, but that the detailed course of the failure depended upon the listening criterion. Piéron (74, p. 302) concluded from the literature that the "point of time" is 10 msec for hearing and about 100 msec for vision.

Rosenzweig and Rosenblith (77) investigated electro-physiological responses of the auditory nervous system, finding that as the interval between monotic clicks was increased to a few milliseconds, the single perceived click began to split into two. They describe it as follows:

> The click seems to be double-humped with the second hump originally small but growing to the size of the first as the interval is increased. When the time interval is about 10 msec, fusion has ended and two separate clicks are heard.

Zwislocki (100), concerned with intensity, found that the threshold of audibility for pairs of clicks decreased as the time interval between the clicks in each pair decreased. Sandel and Stevens (78) demonstrated the trading function between intensity and duration of one noise component embedded in a pair of different noises. The trading relations did not show a simple pattern.

A publication by the School of Aerospace Medicine, Elliott (21), reviews recent research and reports an investigation of the phenomena of backward and forward masking. "Backward masking" denotes the situation in which one auditory stimulus is masked by another auditory stimulus which follows it in time, whereas in "forward masking" the temporal positions of masking and masked signals are reversed. White noise was used in the study as the masking signal and a 1000 cps tone was the masked or probe signal. Masking intervals, the time duration separating the probe signal from the masking signal, were of varying durations from 0 to 50 msec. One of the graphically-presented results of the backward masking phase of the experiment was that the slope from 50 to 15 msec was considerably less than that from 15 to 0 msec, indicating more rapid changes in masked threshold as the masking interval decreased. This time of slope change, not apparent in the forward masking results, was compared favorably with the results of Rosenzweig and Rosenblith (77). It was hypothesized that this result reflected "central masking" processes.

Quite a different approach to successiveness is a group of studies dealing with an auditory phenomenon analogous to visual flicker, auditory flutter. Miller (63, p. 70) discusses the theoretical and instrumental approaches to such studies. Noteworthy in the present frame of reference is the interruption of white noise, Miller and Taylor (67), and the interruption of speech, Miller and Licklider (66). Miller and Taylor reported that the point at which an interrupted noise fuses into a continuous perception depends upon the rate of interruption, the sound-time fraction, and the intensity of the noise. Differential sensitivity to changes in the rate of interruption with a sound-time fraction of 0.5 is poor above 250 interruptions per second. The writers suggested that at 250 bursts per second the temporal separation between bursts is 2 msec, and the last 5 mm of the basilar membrane are active at the moment the next burst begins to travel the cochlear duct.[1] Miller and Licklider observed that if the frequency of interruption of one-syllable words was low enough, the articulation score must be equal to the product of the speech-time fraction and articulation score for uninterrupted speech. It was also observed that between the very low and the very high frequencies of interruption the functions passed through a minimum, a maximum, and then another minimum per cent of intelligibility. In a clinical investigation, Ax and Colley (4) employed auditory flutter as a diagnostic tool for neuropathic and psychopathic patients. They interpreted their results to indicate that temporal acuity thresholds for audition can be measured with sufficient reliability to be useful for the diagnosis of cerebral pathology.

In an experiment designed to determine the accuracy with which subjects could count the number of short 1000 cps tones in a temporal series, Garner (31) reported that the accuracy of counting was a function of both number of tones and repetition rates. The investigator concluded that very low numbers could be accurately counted at repetition rates as high as 12 per sec, but at numbers

[1] This assumption is based upon Békésy's report (G. v. Békésy, *Akustische Zeitschrift*, 8, 66-76 [1943], p. 72) that approximately 5 msec are necessary for a pulse to travel the length of the cochlear duct.

above five or six, counting accuracy decreased with rates of six or more per second. Garner also observed that tone intensity (55 and 95 db SPL) and duration (5 and 40 msec) had no effect on counting accuracy. Cheatham and White (15) also investigated the auditory perception of number, the subject being instructed to report the number of pulses heard. The stimulus employed was a 1000 cps tone, and pulse rates were 10/sec, 15/sec, and 30/sec. They found that consistent slopes were maintained, indicating a range of perception rates of only 9-11 pulses/sec, corresponding to a range of objective rates from 10-30/sec. From this and other numerosity studies White and Cheatham (98) concluded that there is some temporal process in the central nervous system that limits and orders the perceptual events of the major sense modalities.

Callaway and Alexander (13) investigated terminal reaction times to both visual and auditory signals. Relationships between reaction times and impulse frequencies were plotted and non-linear perturbations in reaction time were observed at stimulus frequencies near 10 cps. This was presented as evidence supporting Lindsley's Neuronic Shutter Theory (58), previously mentioned.

Summarizing, studies have shown that temporal resolving power for a variety of auditory stimuli is not solely dependent upon time. As Peterson (73) has observed, "A strict time domain analysis of acoustical signals is obviously just as narrow as any other isolated approach to signal analysis". The acoustic event, its intensity, and its paired partner all play a role in human resolving time. It can be concluded, however, that for clicks, at least, successiveness can be perceived within a range of 2-10 msec. Below the level of successiveness, the paired signals are heard as one rather than two. Finally, there appear to be several reliable methods of approach to the problem of temporal succession. Among these are monaural and binaural auditory stimulation, electrophysical activity measurement, auditory flutter, and numerosity.

TEMPORAL ORDERING AND PATTERNING

In 1959 Hirsh (41) published the results of a series of experiments utilizing various combinations of frequencies and durations of tones, clicks, and noises. His purpose was to determine how great a temporal interval must intervene between the onsets of a pair of sounds in order for the listener to be able to report which of the two elements came first. He pointed out that intervals of a few milliseconds are sufficient to separate two brief sounds so that the listener will report there are two rather than one. Hirsh employed a turntable with peripherally-mounted cylinders to engage a series of microswitches. These microswitches were connected to both channels of an electronic switch which triggered the onsets of the two sound stimuli. By this method he obtained sounds with onset differentials of 0 to 60 msec, usually in 20-msec steps. The stimuli were terminated synchronously, providing a paired stimulus of 500 msec in most instances. Five trained subjects were presented the stimulus pairs according to the method of constant stimuli with forced choices. Hirsh concluded from his results that a separation time of between 15 and 20 msec is required for the listener to report correctly (75% of the time) which of two sounds precedes the other, and that this minimum temporal interval appears to be independent of the kinds of sounds used. Hirsh suggested that the length of the required temporal interval may reveal that judgment of order requires other mechanisms than those associated with the peripheral auditory system.

Small and Campbell (83) conducted an investigation in which one of the variables manipulated was the duration of the interval between cessation of the constant stimulus and the onset of the variable stimulus. The results were reported in terms of the Weber ratio, $\triangle T/T$, where $\triangle T$ was the increment necessary to give criterion performance when added to a "base" duration, T. The effect on $\triangle T/T$ of increasing inter-stimulus intervals from 50 to 3200 msec was found to be independent of the frequency components of the stimuli.

The effect of a loudness cue on the perception of the temporal order of dichotically presented all-pass equal intensity clicks was investigated by Babkoff and Sutton (6). As the clicks were judged more often as equal in loudness, the percent correct judgment of temporal order decreased. It was concluded that the interaction between dichotic stimuli resulting in loudness changes provides cues, which for part of the temporal continuum reverse the direction of the relationship between separation and correct judgment of temporal order.

Broadbent and Ladefoged (11) indicated that the Hirsh findings (41) apply under the optimal condition that the pair of sounds is presented repeatedly until the listener is prepared to judge. They also suggested that when the temporal interval between two sounds that are already close together in time is changed, the overall quality of the two-sound complex is changed. These authors spliced varied non-verbal recorded pairs of a *pip*, a *hiss*, and a *buzz*, using two duration combinations, 120 msec and 30 msec. Initial results revealed that discrimination of the 150-msec stimulus pairs was in one case perfect (pip/buzz) and at chance level for other combinations. After repeated exposure, however, the threshold fell to the magnitude found by Hirsh (41). The investigators believed that the sounds were discriminated on the basis of differences in quality and not by a difference in perceived order. They concluded that their findings were evidence for a perceptual mechanism working on discrete samples of sensory information rather than on continuous flow. They concluded:

> It appears that when two stimuli fall in the same sample, their order is not immediately apparent perceptually ... listeners have to be trained to interpret as order those cues which the ear transmits about the relative arrival of the stimuli.

They suggested that Hirsh's data revealed the reliability of these latter cues whereas their own results had more bearing on the minimum length of the perceptual sample below which training is necessary (either by giving repeated experience, or, in the case of speech, by teaching subjects to discriminate similar differences in sounds) before differences in order can be appreciated.

It should be noted that in addition to the training factor, the Hirsh (41) and Broadbent and Ladefoged (11) studies were somewhat different in design. Hirsh measured *onset* differentials completed by a mixture of the two elements with eventual simultaneous termination. Broadbent and Ladefoged, however, measured onset differentials with the added cue of complete termination of one element before the onset of the second element. This difference in design may be thought of as overlapping versus successive events. The studies differed significantly in design (and results), and in both are found generalizations to speech perception from nonverbal data.

Researchers at the Haskins Laboratories have been studying the cues on which speech perception depends through the use of techniques which synthesize and modify the sounds of speech. Simplified hand-painted spectrograms based on sound spectrographic displays are converted to sound by a machine called a pattern playback. A wide range of experimental changes thus can be made in the painted spectrograms on the basis of listening trials, quantizing in terms of the frequency, intensity, and band widths of the major energy peaks, the formants. One such study, Liberman et al. (52), emphasized the role of learning in the perception of discrimination peaks. Discrimination peaks develop at phoneme boundaries, according to the investigators, when perception becomes discontinuous or categorial as a result of articulatory changes such as tongue movement for /d/. But when acoustic cues are produced by movements that vary continuously from one articulation position to another, perception tends to change continuously and there are no peaks at the phoneme boundaries. To test the role of learning in this process, various durations of silence were introduced as part of a synthetic speech pattern and in a non-speech context. In the speech case the durations of silence separated the two syllables of a synthesized word, causing it to be heard as *rabid* when the intersyllabic silence was of short duration and as *rapid* when it was long. The boundary between /b/ and /p/ lay at about 70 msec of the silent interval according to pooled responses of 12 subjects. With acoustic differences equal discrimination was more acute across the /b, p/ phoneme boundary than within either phoneme category. Liberman et al concluded, "This effect approximated what one would expect on the extreme assumption that listeners could hear these sounds only as phonemes, and could discriminate no other differences among them ..."

Another study by the Haskins group, (53), used synthetic approximations of the syllables /do/ and /to/, which differed only in the relative time of onset of the first and second formants, the first formant being "cut back" in 10-msec steps from 0 to 60 msec. Measures of both the location of the phoneme boundary on the scale of relative onset-times and the discrimination ability were obtained from naive subjects. In addition, scores were obtained from a set of non-speech control patterns produced by inverting the speech patterns on the frequency scale, thus preserving the temporal variable while precluding phoneme recognition. Once again it was found that discrimination was considerably better across a phoneme boundary than in the middle of a phoneme category. The discrimination ability on the inverted controls was poorer than for the speech stimuli and revealed no increase at stimulus values in the region corresponding to the location of the phoneme boundary. In discussing their results, Liberman et al noted that for speech stimuli straddling phoneme boundaries (those predicted from phoneme labeling results to yield the best discrimination), the difference in time of onset required to give 75% correct judgments was slightly less than 12 msec. This difference in time of onset was compared with the results of Hirsh (41) in which the 75% point for pure tones was 17 msec. The superiority of discrimination for speech patterns was offered as support of the assumption that the discrimination peak at the phoneme boundary is a result of learning.

Chistovich (17) studied discrimination of a complex signal obtained when the sounds that made up the signal were cut in simultaneously and when the component sounds were cut in with a time lag. The results showed that depending upon how the complex sounds were cut in they were either received as one signal /ɛ/ or were separated into two independent signals /i/, /ɑ/. When the sounds were cut in simultaneously they were heard as one signal with precise timbre determination. The segregation of one of the component complex sounds from the aggregate perturbation proved to be "extremely difficult in the case where sounds were cut in simultaneously".

In what might be termed a follow-up study to Hirsh (41), Hirsh and Sherrick (42) investigated the hypothesis that a single quantitative process is shared by all three sense modalities, visual, auditory, and tactual. Their auditory experiment employed 0.1-msec *bings* and *bongs* separated by 10, 20, and 30 msec with the bing leading one-half the time and the bong leading the other half. The results showed that the amount of time that must intervene between two events in order for the subject to report correctly (75% of the time) which of these two events preceded the other was approximately 20 msec. The function was seen to be essentially the same if the events were either two sounds of different quality, or two light flashes in different places, or two vibrations, one to each hand, or two sounds, one to each ear, or two different stimuli, one to each of two sense modalities. Hirsh and Sherrick maintained that Broadbent and Ladefoged's (11) explanation of "quality judgments of the two-sound complex" in the previous Hirsh study (41) was not appropriate to their present study.

Using quite a different approach, Ladefoged and Broadbent (49) played to their subjects a series of tape-recorded sentences with an extraneous sound superimposed on the recording. The listener was to indicate the exact point in the sentence at which this sound (an /s/ in some cases, a click in others) occurred. It was found that errors were made which were large compared with the duration of a single speech sound, suggesting, they believed, that the listener does not deal with each sound sepatarely but rather with a group of sounds. Errors were reduced if the sentence consisted of a series of digits or if the listeners were trained in phonetics. The direction of error was usually to refer the extra sound to a point earlier in the sentence than it actually occurred. Black (9) investigated another aspect of attention in a study of repeated and overlapping speech patterns. Recorded nonsense syllables were delivered to the headsets of listeners via two channels, a direct line and a line with controlled amounts of delay from 0.00 to 0.33 sec. The amount of delay was varied in 0.03-sec increments and the effect of this variable upon reception of the syllables was measured. Black reasoned that the listener who focused on the original signal was likely to hear the initial consonant correctly as he tried to hear three sounds in their correct temporal sequence, and the listener who focused on the delayed signal may have heard the final consonant correctly. He suggested that either preparatory set was possible, somewhat like the optical response to a reversible plane figure.

Kinney (48) investigated patterns, both auditory and visual. The effects of pattern type and element combination upon temporal discrimination in auditory patterns and upon spatial discrimination in visual patterns were measured. In the auditory portion the output of a pattern playback with painted spectrograms was taped and delivered to a loudspeaker system. The investigator separated two different tones of 50 msec duration by 50 msec of silence for the standard condition. For the variable conditions, the silence was of 10, 20, 30, or 40-msec durations. Two degrees of frequency separation were used, a near-condition (1200/1440 cps) and a far-condition (1200/3480 cps). Comparison judgments at each temporal displacement were made by 138 subjects in small groups. The resultant curves started with a chance level of discrimination at 10 msec and for all patterns (pyramid, triad, and ascending) and showed increasingly larger percentages of correct judgment as the length of the temporal displacement was increased. The liminal values for both frequency curves at the 75% point were 30-40 msec. Whenever the results were above chance level, the near-condition yielded better discrimination than the far in both the auditory and visual (spatial) discrimination tasks. Kinney concluded that for both audition and vision, temporal discrimination varied with the type of pattern in which the discrimination was imbedded, and that discrimination was better when the elements in the pattern were of relatively near frequencies than when they were widely separated. One repetition of the basic elements improved discrimination, but there was no further increase with further repetitions.

Another study of auditory patterns, Heise and Miller (37), showed that settings of a tone within a pattern were influenced not only by the tones that preceded it but also by tones that followed.

Warren and Gregory (93) reported yet another method of investigation of temporal perception in speech. Using recorded tape loops with repetitions of words or phrases without pause, they found that the verbal organization undergoes abrupt transitions into other words or phrases, sometimes accompanied by apparent changes in component sounds. The word *rest* was found to shift to *tress*, *stress*, or even to *Esther*. They compared this verbal alteration to the principle of the reversal of visual figures.

In a recent clinical study Efrom (20), building upon his previous experiments designed to support the hypothesis that certain information (relative to the time of occurrence) is transferred from non-dominant to dominant brain hemisphere before temporal discrimination of simultaneity and order can be performed, examined visual and auditory temporal perception of aphasic patients. Eleven aphasic subjects with unequivocal lesions of the left hemisphere were matched with 5 non-aphasic patients who served as controls. The aphasic group was further divided into predominantly "expressive" and predominantly "receptive" sub-groups. The auditory stimuli in his study consisted of tones of 10 msec duration with two pulse frequencies 2500 pps (*bleep*) and 250 pps (*bop*). No attempt was made to make the sounds of equal energy, or of equal apparent loudness. The subject reported which sound was first when the pair of sounds was repeated every four seconds. The 75 per cent threshold values for determination of correct sequences were 75 msec for the control group, 400 msec for the expressive aphasics and 140 msec for the receptive aphasics. Efrom compares auditory and visual results favorably with those of Hirsh and Sherrick (43) and accounts for the deviation of his controls (75 msec versus 20 msec for Hirsh and Sherrick) on the generally poor performance on tasks of sensory discrimination by hospitalized patients. Efrom concluded (1) that lesions which produce disturbances in the discrimination of temporal sequence are all in the hemisphere which is dominant for speech functions and (2) that every subject with a dominant hemisphere lesion who had difficulty with temporal analysis also had some degree of aphasia.

Before concluding this section, it should be noted that a factor of *forward dependency* is common to several theories and a number of investigations. Results in backward masking (21), auditory patterning (37), and placement of an extraneous sound in a sentence (49) illustrate the influence of future events upon the present. These studies provide support for the theories of Huggins (44) who postulated the present as always immediately in front of us, and of Lashley (50, chap. 4) who wrote that sequential order is pre-imposed. Finally, Licklider (55) emphasized the important aspect that the incoming messages arrive in such an order that appropriate comparison patterns can be activated ahead of time.

By way of summary, ordering and patterning of acoustic events have been studied from both the verbal and non-verbal approaches and with both normal and abnormal listeners. Valuable, though to some extent conflicting data have been obtained on the ordering-time requirements of acoustic and synthetic speech samples, the environmental patterning influences, and the roles played by acoustic parameters other than the primary variable, time. A common finding in several of the studies, forward dependency, was pointed out. A possible linguistic application of this forward factor will be discussed in the two succeeding sections.

LINGUISTIC ASPECTS

In addition to the contributions made by linguists in the areas of segmentation and phoneme sequence, the specialized study of linguistic change offers some explanation for normal and abnormal ordering of phonemes in time.

In 1895, according to Sturtevant (86, p. 32), Meringer and Mayer published the first list of errors of phoneme assimilation under the title *Versprechen und Verlessen.* The first English collection was Bawden's *A Study of Lapses* published in 1900. From these publications Sturtevant analyzed and enlarged the classifications of linguistic pronunciation errors. One such classification was described as follows (86, p. 50), "When two sounds or groups of sounds interfere with each other, the result is sometimes an exchange of places, a process which has long been called *metathesis* in our grammars." Sturtevant observed that most of Meringer's and Bawden's examples were metathesis of sounds which were not immediately contiguous. He showed, however, that two consonants in contact may suffer metathesis as with Bawden's *wist* (*wits*), Anglo-Saxon *axian* (*ascian*), Modern English *ax* (*ask*), Chaucer's *brid* (*bird*), and many Greek words. Sturtevant suggested that cases of associative interference that are not repeated or imitated usually concern sounds which stand under similar accentual conditions whereas cases that become permanent features of a language almost as regularly concern neighboring sounds. He explained (86, p. 52):

> Since an accentual group commonly embraces a whole word or several words, an interference between sounds of similar accentual relations usually involves two or more words, and therefore cannot recur until the same combination of words occurs again. An interference between

neighboring sounds, on the other hand, usually involves only one word, and may therefore recur whenever the word is spoken. So, although changes of the second sort are far less common than the others, each one of them is more likely to be repeated and hence to find imitators.

In a more recent publication, Sturtevant (87, p. 92), explains that Spoonerisms, for example, "sew her to a sheet", are so grotesque that they are sure to be noticed, and no one would imitate them except in jest. Sturtevant concludes that these utterances could have scarcely any influence on the development of a language.

Goldberg (32, pp. 206-207) has shown effects of metathesis among languages. He gives examples such as Latin *periculum* producing Spanish *peligro*, and Latin *crocodilus* (crocodile) to Middle English *cockadrill* and Spanish *cocodrillo*. Goldberg observes that as in the pronunciation of children, certain errors of adult speech sometimes point the way to future linguistic developments.

Gray and Wise (33, p. 372) describe common American-English examples of metathesis as [s] plus a plosive (*waps/wasp*, *gaps/gasp*) and [r] preceded or followed by a consonant (*purty/pretty*, *patteren/pattern*). These authors classify *ax/ask* as Louisiana French-English.

Several types of metathesis are given by Sturtevant (87, pp. 89-93) together with associated, though non-transpositional, phenomena of assimilation. The metathesis categories are (a) anticipation with loss (*hemlet/helmet*); (b) distant metathesis (*evelate/elevate*) or Spoonerism (*I fool so feelish/I feel so foolish*); (c) more complicated metathesis (*presumably prular/plural* – this alteration was described as r...l...l...r...l to r...l...r...l...r); and (d) lag with loss – only one certain example – (*by help sending/by helping send*). In addition, anticipation alone (*phortograph/photograph*), anticipation with substitution (*skea and sky/sea and sky*), and anticipatory loss (*Poasties/Post Toasties*) were reported. It is interesting to observe that all of these errors except "lag with loss" are anticipatory in nature, borrowing a sound or sounds from some forward or future element. The role of time would appear to be of major significance in these errors.

Summarizing, metathesis, the exchange of position of speech sounds, is a linguistic phenomenon dating back to ancient languages, yet common today. The transposed sounds may be immediately contiguous or separated by other speech sounds. Most examples are of an anticipatory character rather than a lag, and some phonemes seem to be more likely metathesized in certain sound environments than others. Generally, such phonemic anachronisms are classed simply as mispronunciations or slips of the tongue, but often repeated performance and imitation result in linguistic change. The reasons for the universality of the metathesis phenomenon are not considered by the linguists. The suggestion is strong, nevertheless, that language becomes molded by laws of economy of energy and the natural incapabilities of the speakers. The basis for these incapabilities is a matter of some speculation, but clinical observations, it will be seen, offer some information on the general nature of the problem.

CLINICAL CONSIDERATIONS

Just as the study of medical pathology has contributed significantly to the understanding of the human body, so too, it is believed, the study of speech and hearing pathology offers possibilities for better understanding of the normal speech act. This section is concerned with a breakdown of the time patterns of speech as reflected in temporal sequences, or a type of clinical metathesis. To the writer's knowledge, however, this is the first use made of the term, metathesis, to describe a clinical symptom.

Sound sequence confusion has long been recognized as a factor in early normal language development or baby talk. Van Riper and Irwin (91, p. 29) have pointed out that young children, when mastering pronunciation skills, frequently show reversals, insertions, and omissions of sounds. They cite *pasghetti* as an example, but there are many other illustrations such as *aminal*, *swikers*, and *hippopotasum* which are often short-lived examples of the same metathesis or transposition of sounds described in the previous section.

Of more concern, however, is the persisting confusion which seems to stay with some children even into adulthood. In 1937 Orton (72, pp. 144 ff.) described a difficulty in repicturing or rebuilding in the order of presentation, sequences of letters, of sounds, or of units of movements. This symptom he found to be a common factor in a ten-year study of each of the following disorders: developmental alexia, special disability in writing, developmental word deafness, motor speech delay, stuttering, and developmental apraxia.

More recently, many authorities in speech, hearing, and reading have expressed concern for a problem described in a similar fashion, but often diagnosed as simply disorders of auditory memory span (7, 96, 8), association (35), childhood aphasia (69), or articulatory apraxia (70). Berry and Eisenson (8, p. 103) describe a ten-year-old child who, after a year of retraining, still used such expressions as *tapetoes and gavery* for *potatoes and gravy*, *pip it uk* for *pick it up*, and *my dloy* for *my dolly*. The problem has been described by Hardy (35) in these words:

> Most of these children perceive sound quite well but cannot employ what they perceive ... what is most obvious is that they cannot naturally listen, understand, store and recall symbolic structures involving a time order and a stress pattern.

Monsees (68, 69) has made use of the symptom as a diagnostic tool for the deaf and young aphasic. She maintains that a child who fails to learn by the whole-word method involving temporal patterning but succeeds in learning by the elemental approach involving spatial patterning has indicated something significant about the nature of his disability. She concludes (69), "He has indicated that he has a central disorder involving the perception of temporal sequence and we believe that this is the basic nature of the disability of aphasia." Furth (30) has explored Monsees' observations with a recent inconclusive study comparing visual learning of sequences in aphasic children with that of deaf and hearing controls.

In adult aphasia as well, the phenomenon has been observed. As early as 1885 Lichtheim (54) described the phenomenon by quoting a patient's account of his own aphasia: "'Another mode of paramnesia (erroneous use of known and remembered sounds) consisted in inverting the letters of the syllables in a compound word, of which I had recovered possession. Thus for *raisin* I said *sairin*, for *musulman* I felt inclined to *sumulman*'." Fry (29), describing the phonemic substitutions of an aphasic patient, noted that frequently the patient gave the wrong phonemic sequence without apparent awareness. Efrom (20) also eludes to the problem of temporal perceptual difficulties among his experimental population of aphasics.

Morley (70, p. 223) describes consonant transposition with reversal of order as an aspect of articulatory apraxia. She cites such common examples as *breakufs* for *breakfast* or *bilicles* for *bicycles*. West, Ansberry, and Carr (96, p. 224) state that "Usually the only factor that can rationally be connected etiologically with the speech defect (phonetic lapses and transpositions) is shortness of auditory memory span". Backus (7, p. 133) agrees that the defect is one of short auditory memory span. Van Riper and Irwin (91, p. 29) maintain, however, that to realize that the word *nose* has three distinct sounds, [n], [o], and [z], and in that sequence, requires some memory and some recognition of sound characteristics. "But it requires something more – the ability to perceive a temporal sequence and to recognize where each sound belongs in that sequence ... The word *zone* has the same phonemes, but in reverse order." Rosenblith (76, p. 822) suggests that future studies of temporal aspects of sensory communication are likely to make contact with studies of short-term memory.

De Hirsch (38) has analyzed in greater detail the actual perceptual phenomenon and shown how it develops from a speech disorder into a total language disorder.

In order to read a little word like 'mat', a sequence of letters seen, a sequence in space has to be translated back into a sequence of sounds heard, a sequence in time ... Most children who later on develop reading disabilities seem to have trouble with patterning the units of words and sentences into spoken speech ... The same boy who says 'criceripsies' for 'rice crispies' is the one who later reads 'was' for 'saw' and 'now' for 'won'. Reversals and confusions in the order of sequences ... are generally not confined to syllables and words.

Since 1960 a multi-phased investigation has been published under the title of *Studies in Tachyphemia*. *Tachyphemia*, also termed *cluttering*, is defined by Arnold (3) as a disability of language formulation resulting in confused, hurried, and slurred diction on the basis of a congenital, inheritable and constitutional limitation of the total psychosomatic personality structure. Arnold points out such characteristics as (a) inversions of sounds, syllables, words and sentence particles reflected in deficient auditory feedback, and (b) disorientation with elements of time and space (rhythm, rate, syntactic order). Shepherd (81) observed transpositions of sounds in phonetic transcriptions of several typical examples of cluttered speech.

Summarizing, the linguistic term, metathesis, has been introduced to describe a phenomenon observed in normal language development and a clinical symptom apparent in some cases of abnormal language development, childhood brain damage, and adult aphasia. In its simplest form this phenomenon is reflected in the transposition of sounds within a word or group of words. Like linguistic metathesis, the transpositions described are of the anticipatory type, borrowing a phoneme from the future utterance to trade for the presently-emitted speech sound. This trading function in time, as observed in the theoretical, experimental, and linguistic sections, would appear to be based upon the manner in which the human system deals perceptually with the incoming speech signal. It seems logical that an explanation of mere articulatory disability is insufficient to describe the confusion since the phonemes per se do not suffer from faulty production, but rather are rearranged in time. The etiology of the symptom is a matter of some disagreement, but has been associated with aphasia, auditory memory span, disturbance in the ability to perceive temporal order, and the often sub-clinical tachyphemia. Its occurrence in speech has been associated with subsequent problems in reading, spelling, and typing. In conclusion, there appears to be a clinical metathesis in language disorders.

PART II

AN INVESTIGATION

PROBLEM AND PROCEDURE

Though considerable information on temporal aspects of audition is available, there is a decided lack of data obtained through the employment of true verbal stimuli. If, as reported in Part I, language perception is in fact highly dependent upon the temporal sequence of the component elements or events, then it seems that data on temporal ordering of verbal events are essential to the understanding of the perceptual act and its breakdown. This investigation was designed to ascertain the onset differentials required for an individual to resolve monaurally the temporal order of various pairings of recorded voiced consonant phonemes. It was hoped also that information might be obtained to classify in terms of temporal resolution abilities, (1) differences among phonemes or phonemic classifications, (2) differences in the relative positions (first or second) of a member of a phonemic pair, and (3) differences between selected phonemes and two similarly-paired pure tones.

The following procedural steps were employed in conducting the experiment:

A. Carrying Out of Pilot Investigations.
B. Composition of Stimulus Tapes.
    1. Recording.
    2. Onset-Timing.
C. Composition of Master Tapes.
D. Testing of Subjects.

Pilot Study I was conducted in an effort to determine if a technique could be developed which would allow for extremely accurate manipulation in the time domain of the respective onsets for two tape-recorded signals. In addition, this pilot investigation was designed to reveal the feasibility of the overall experimental procedure and to discover the type and range of data to be expected from the listening judgments. On the basis of the results of this preliminary study it was concluded that the general experimental technique employed was satisfactory and could, with some modifications, be used as the procedure for the major investigation. Moreover, certain specific stimulus pairs of phonemes were shown to be more sensitive to temporal factors, whereas others proved to be relatively insensitive.

Based on this latter finding, Pilot Study II was carried out to collect data which would aid in the selection of the specific phoneme pairs that would prove most sensitive to the temporal restrictions to be imposed in the major investigation.

Inasmuch as these preliminary pilot investigations served as the bases for the major study, a brief account of the two studies is provided in Appendix A. It should be noted, however, that important findings of the studies which helped determine the specific procedures of the major investigation are pointed out, when applicable, in the succeeding pages of this section.

DESIGN OF THE EXPERIMENTAL SEQUENCE

The major investigation was designed to make use of four master tapes, each of which provided the experimental conditions for three different phoneme pairs plus a common pure-tone pair. (The pure-tone pair, 1200/250 cps, was used as the final condition of each experimental sequence in order to obtain a constant performance criterion over all subjects. In addition, this condition provided comparison data with one of the experiments of Hirsh [41] and with the first pilot investigation). The four tapes, therefore, accounted for the 12 phoneme pairs selected on the basis of Pilot Study II as well as one pure-tone pair.

Each tape was played to 12 different subjects. This procedure allowed for a variety of possible orderings. The first four subjects received the three phoneme pairs in A-B-C order with "D" symbolizing the final pure-tone pair. (Inasmuch as this "D" condition was the only condition common to all experimental sessions, it was believed that it should be appended to each instance to avoid interference with the ordering of the primary conditions, the phoneme pairs). The second group of four subjects received the taped stimuli in the sequence of B-C-A-D, and the third group was given the sequence of C-B-A-D. This procedure was followed for all blocks of 12 subjects, a different tape containing different stimulus pairs being used for each. Thus a total sample population of 48 was required. The order of occurrence of the 12 phoneme pairs was randomized by a blind drawing procedure (without replacement) from a box containing thoroughly-mixed identically-cut cards, one each for the 12 pairs.

The experimental conditions applied to each stimulus pair were based on the design requirement of four presentations of each onset differential. (As pointed out in the results of Pilot Study I, Appendix A, ten-trial results yielded only slightly better results than the first trial of each presentation. It was therefore decided that four exposures to a given condition would be sufficient for collection of reliable data.) The times selected for the amounts of onset disparity were –70, –50, –30, –10, 0, 10, 30, 50, and 70 msec. (It should be noted here that the minus sign preceding a disparity time indicates simply that the second phoneme of a pair would lead the first. This is contrasted with the onset disparity shown as a positive value, in which the first phoneme would lead the second by the given number of milliseconds. Thus, for the v/l pair, 50 would indicate that /v/ led /l/ by 50 msec, whereas –30 would indicate that /l/ led /v/ by 30 msec.) The selections from –70 to 70 msec represent a somewhat broader range than those used in the first pilot study; they were chosen for the reason that both the upper and lower limits of discrimination had not been obtained in the pilot.

Four successive randomizations of the nine temporal conditions were performed for each of the four stimulus pairs of each of the four master tapes. Again the randomizing process was accomplished by the blind drawing method previously described. The result was 36 experimental presentations for each stimulus pair. In addition, prefixed to each 36-item sequence was a group of four presentations of the stimulus pair with an onset disparity of 80 msec. This four-item group was provided to acquaint the listener (a) with a sample of the auditory experience of a given stimulus pair, and (b) with the task to be performed. This feature thus was to serve as a brief introduction although this fact was not to be disclosed to the subjects nor were the subjects' responses to the first four items of each stimulus pair included in the data tabulation of the experimental conditions. There was no "break" in the experimental routine between the fourth and fifth presentations; thus the prefixed stimuli were continuous with the true experimental conditions. (Because there was no possible correct ordering of the zero-onset-disparity conditions, these data were also not tabulated with the body of the experimental scores, but were in every other respect handled as experimental conditions.)

In summary, four master tapes were prepared, each of which contained 40 temporal conditions for each of four stimulus pairs (three phoneme pairs and a pure-tone pair). Each tape was presented to three groups of four listeners each, making a sample population of 48. The ordered sequence of presentation of each four-pair tape was changed for each group of four listeners. The design thus provided 48 judgments for each temporal condition (four subjects each making 12 judgments).

COMPOSITION OF STIMULUS TAPES

Recording

On the basis of the results of Pilot Study II, 2.0– second sustained utterances of /v/, /ð/, /z/, /r/, /l/, /m/, and /n/ were recorded on magnetic tape. The experimenter, who speaks in the General American dialect and has no history of a speech disorder, served as the speaker. All consonants were produced at a constant zero

VU reading with a constant lip-to-microphone distance.[1] In order to ensure that the recorded phonemes were valid examples, several expert judges, graduate students in speech and hearing science at the Purdue Clinic, were asked individually to listen to and evaluate the recorded utterances. The pooled judgments indicated that each example was a satisfactorily produced sample of the desired phoneme.

In addition to the vocal stimuli, 2.0–second recordings of a 1200 cps and 250 cps tone were made from the output of a Hewlett-Packard 200CD oscillator. In all recording and playback procedures for the experiment a tape speed of 7½ ips was used to accommodate to the single-speed characteristics of one of the tape recorders employed, the Ampex 601. Recording was accomplished in a two-room recording suite using the instruments listed below. (Detailed descriptions, and measured performance specifications for the major instrumentation used in all phases of the experiment may be found in Appendix B. Included in this appendix are the frequency response characteristics of all equipment used with the exception of the microphone, for which characteristics could not be obtained.) All equipment was found to be ± 3 db with the exception of the headphones, which showed a drop in sensitivity at 4000 cps and higher. The stimulus recording equipment was as follows:

a. Microphone, Altec 21-B;
b. Microphone power supply, Altec P518-A;
c. Preamplifier, Altec 428-B;
d. Line amplifiers, Altec 429-B;
e. Tape recorder, Ampex 300.

An apparently steady-state segment of each stimulus, as determined by monitored VU activity, was extracted from the mid-portion of each 2.0–second recording by removal of a 4¼-in. (567 msec) tape section. Each segment was then spliced into a "clean" tape,

[1] Tiffany (88) found that this method of controlling intensity, when employed for recording sustained vowels, was a reliable procedure for producing stimulus material for the testing of hearing.

allowing several feet between sound samples. For this procedure a diagonal (45 degree) splicing cut was used exclusively, thus creating a standard rise time of 10 msec.

The next step in the procedure was to compose sound stimulus pairs from this recorded supply of single sound samples. The pairs to be created were v/l, v/r, ð/r, z/l, z/r, r/l, m/n, z/n, n/ð, n/r, m/r, ð/l, and the pure-tone pair, 1200/250 cps. The instrumentation used for this procedure was the following:

a. Tape recorder, Ampex 601 and
b. Tape recorder, Ampex 300 (both channels).

The first element of each of the stimulus pairs was transferred from playback of the Ampex 601 recorder to channel #1 of the Ampex 300 recorder. An interstimulus distance of about three yards was allowed. Then, the onset of the stimulus was marked on (a) the Ampex 300 channel #1 recording and (b) the second member of the phoneme pair now ready for playback on the Ampex 601. The marked tape of the Ampex 601 was then reversed a few inches to allow for the prior onset of the channel #1 stimulus in relation to the channel #2 stimulus onset on the dual-channel tape. With the aid of an assistant, both recorders were then simultaneously engaged. The result was a two-channel recording with signal onsets very close, but with channel #1 beginning slightly sooner. This procedure was repeated for each of the stimulus pairs. The resultant dual-channel recorded tape was then cut to form 37-in. tape loops, a loop for each stimulus pair.

Onset-Timing

The next undertaking was to ascertain the exact number of milliseconds separating the members of each stimulus pair, thereby providing a base time (simultaneous onset) from which experimentally-imposed temporal deviations could be effected. The piece of apparatus used for this purpose was the Echo-Vox Sr.[2] This instrument, hereafter termed simply "Echo-Vox", is designed

[2] Kay Electric Company, Pine Brook, New Jersey.

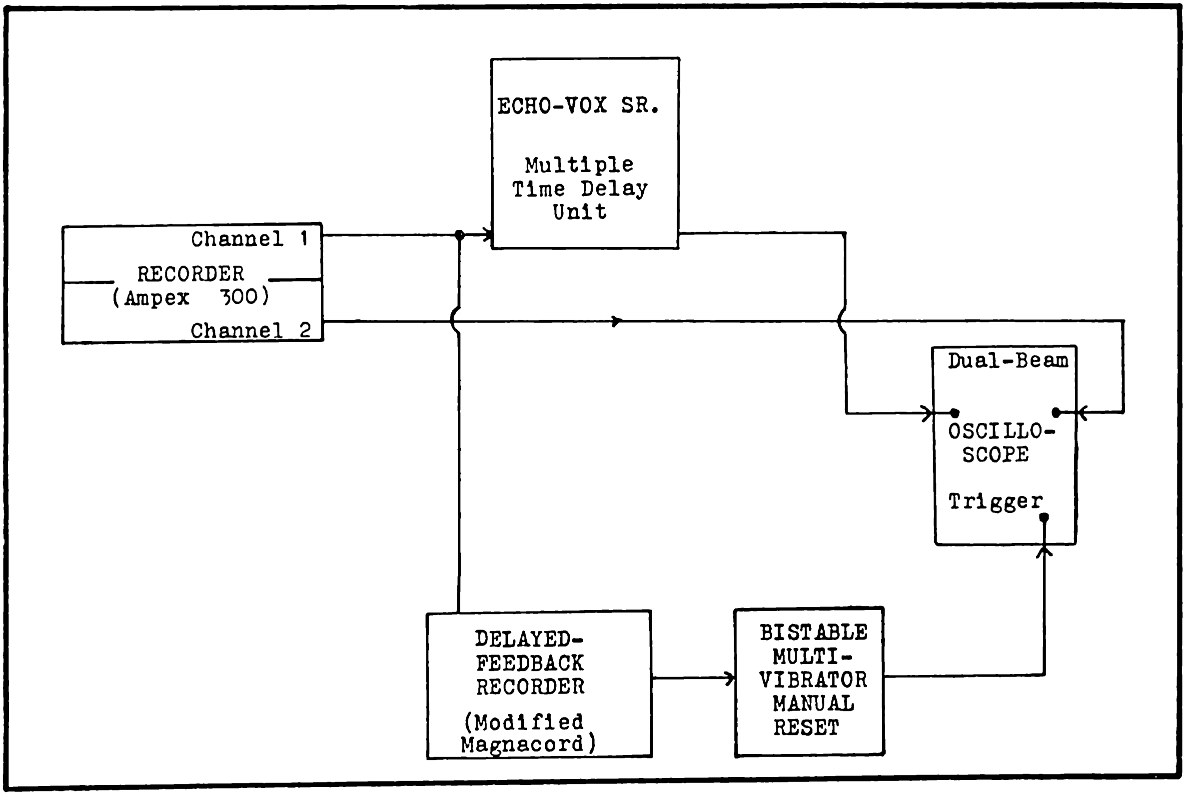


Figure 2. Block diagram of the apparatus employed to effect onset time differentials of paired stimuli.

to introduce controlled amounts of delay into an audio system. It is capable of providing three separate periods of delay within either of two ranges and possesses, in addition, a single, direct channel. The Echo-Vox comprises a recording amplifier, a motor-driven magnetic turntable, a three-channel preamplifier, and a final power amplifier. The signal in use is delivered through two stages of amplification and recorded on the magnetic turntable. Each of three independently movable reproduce heads may be set to any point in a 230-degree arc to produce the desired delays. Time delay ranges of 20-540 msec and 60-1600 msec may be continuously variable throughout the entire range. The other apparatus employed was as follows:

a. Tape recorder, Ampex 300;
b. Delayed-feedback recorder, "Modified" Magnacord;[3]
c. Bistable multivibrator manual reset;
d. Oscilloscope, Tektronix Type 502 Dual-Beam.

Figure 2 illustrates the instrumental array used for the onset-timing procedure.

From the Ampex 300, channel #1 of the tape loop was delivered to channel #1 of the Echo-Vox and subsequently to channel #1 of the oscilloscope. Channel #2 was delivered directly to channel #2 of the oscilloscope. The step-by-step procedure followed for each stimulus loop was the following:

a. Playback from both channels of the Ampex 300 was delivered to the oscilloscope, which was monitored by a technical assistant.

b. The experimenter then gradually adjusted the Magnacord delay control in order to trigger the oscilloscope to begin its sweep of the signal at the desired time. (The bistable multivibrator unit was operated by the assistant to obtain a single sweep display of the oscilloscope.)

c. When the sweep of both channels was in evidence, the signal of channel #1 was delayed by the experimenter through minute

[3] The standard Magnacord PT6-A Tape Recorder-Reproducer, custom modified for laboratory use in delayed sidetone experiments.

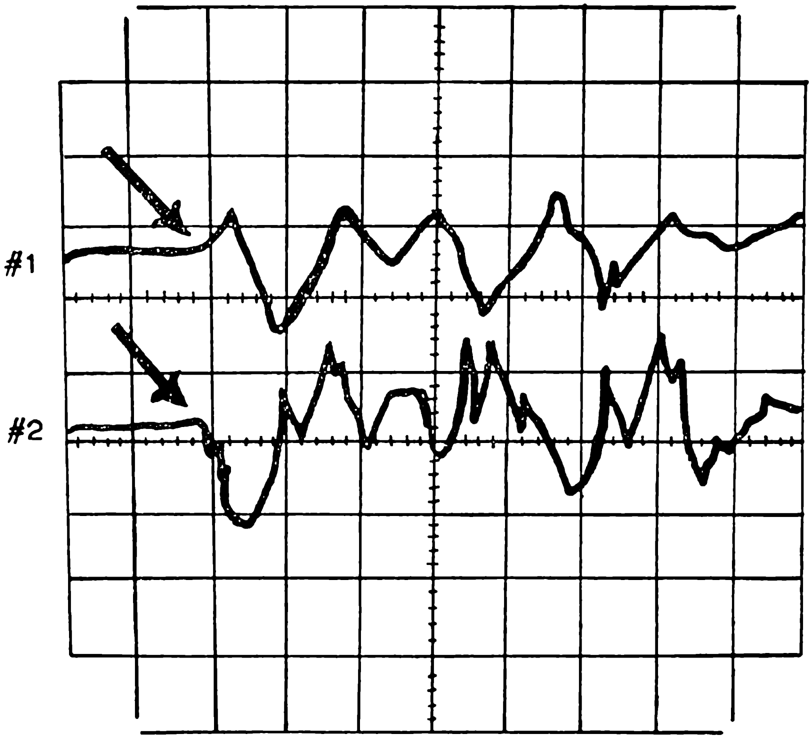


Figure 3. Dual oscilloscopic display on linear time base comparing two waveforms. This display shows signals at the concurrent-onset point (arrows). Each square represents a horizontal time base of 2.0 msec.

adjustments of the Echo-Vox control until it was agreed that both channels were beginning at the same point on the face of the oscilloscope. The delay of channel #1 revealed by the position of the adjustment knob on the Echo-Vox time scale at this point was taken to be zero-onset-disparity (or base time), and this figure of simultaneous onset was recorded for future reference.

The oscilloscopic observations were made using a sweep rate of 2.0 msec/cm. allowing for very accurate onset determinations. Figure 3 illustrates an oscilloscopic display showing a simultaneous onset of sound waves from both channels.

By this method it was found, for example, that when /v/ was delivered to the Echo-Vox and the /r/ signal by-passed the instrument, a delay setting of 313 msec on the Echo-Vox would result in the simultaneous onset of two phonemes when they were once again united.

COMPOSITION OF MASTER TAPES

The instrumentation used for recording the master tapes consisted of the following pieces:

a. Tape recorder, Ampex 300;
b. Echo-Vox;
c. Balancing transformer, United Transformer Co. Model LS-33;
d. Mixer unit, four 600 ohm resistors in a diamond arrangement;
e. Tape recorder, Ampex 601.

A schematic diagram of this instrumental array is shown in Figure 4. Very briefly, the procedure consisted of playing the two channels of a stimulus tape loop into a mixer unit, the first channel going to the Echo-Vox where experimental conditions were effected, and the second channel going to the mixer without alteration. (A balancing transformer unit was used on channel #2 to allow for the ungrounding of one electrical lead prior to entering the mixer unit.) The mixed signal was then recorded on the single-channel Ampex 601.

The experimenter, working with an assistant, recorded the master tape according to the pre-designed experimental presentation order. The steps involved in this procedure were as follows:

a. The playback level of each channel was individually equalized so that a zero VU-level was shown on the meter of the recording Ampex 601 indicating, therefore, a mixed signal with half the represented voltage emanating from channel #1 and half from channel #2;

b. Continuous playback of the stimulus loop was begun by engaging the Ampex 300;

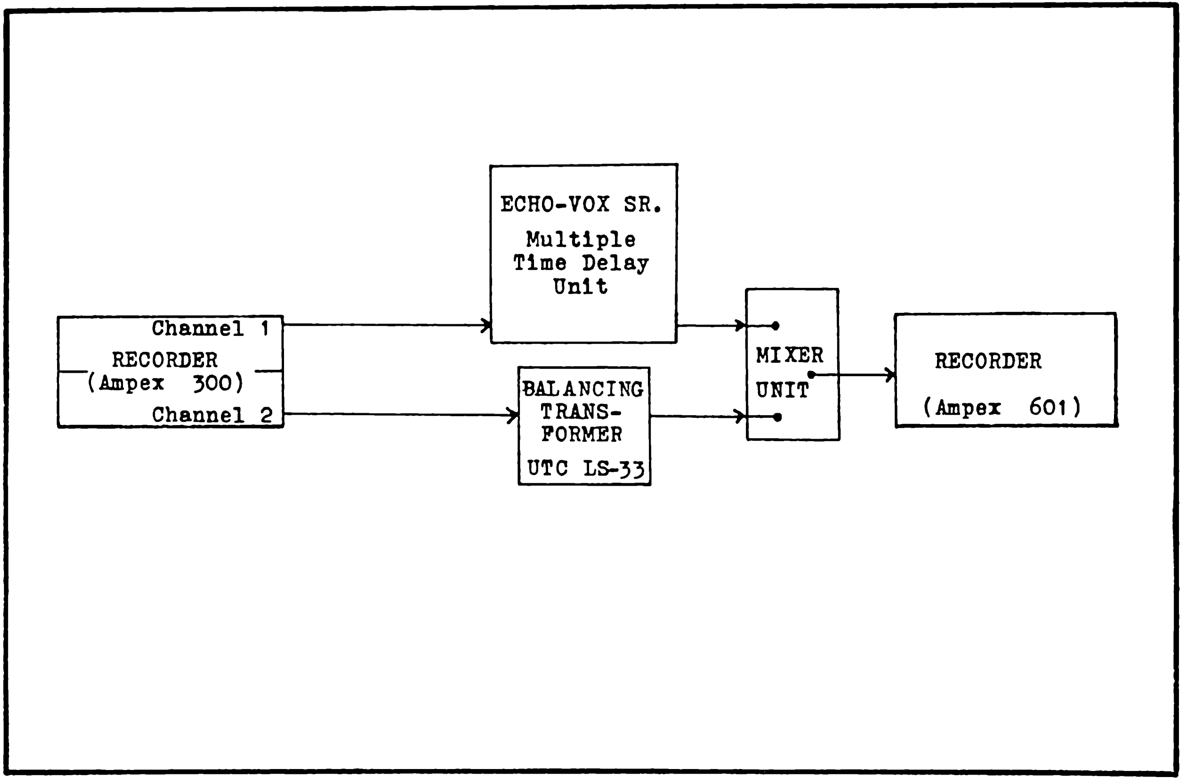


Figure 4. Block diagram of the apparatus employed to make the single-channel master recordings from the dual-channel stimulus tape loops. Temporal displacements were effected at the Echo-Vox Sr., Multiple Time Delay Unit.

c. Simultaneous onset of the two channels was established. To do this a cardboard template with pre-marked time scale duplicating that of the Echo-Vox but indicating only the time settings of concern was used. The "zero" of this template was located at the base time, for example 313 msec for the v/r pair.

d. The Echo-Vox time adjustment control was then moved to the first experimental condition, for example –50 msec.

e. An assistant engaged the recording Ampex 601 immediately following the appearance of activity on the VU meter of that recorder.

f. Once begun, this procedure continued for all conditions of the stimulus pair, the assistant announcing the forthcoming time adjustment at the time that one signal was being recorded. The 5-second interstimulus interval of the ongoing tape loop allowed sufficient time for the experimenter to make the Echo-Vox adjustment necessary for a changed onset differential.

Hirsh (41) reported that equal termination of mixed signals was necessary in order to prevent listeners from using terminal cues as an aid to ascertaining onset order. In the present study the method employed to terminate the now-mixed stimulus pairs was by cutting and removing a tape segment at the end of each recorded signal. After each stimulus onset was marked with ink at the commencement of audible and VU activity from the exposed playback head of the Ampex 601, a 45-degree cut was made after allowing 3 3/4 in. (500 msec) of signal beyond the ink marker. Thus in every case the longer of the two members of the stimulus pair was 500 msec, while the shorter, delayed-onset member could be as long as 500 msec or as short as 420 msec. The following 3 3/4-in. segment was then removed thereby ensuring abrupt and complete termination at the 500-msec point. The tape ends were then spliced, making a new interstimulus separation of 4.5 seconds. A schematic diagram of a typical experimental sequence of the paired stimuli is shown in Figure 5.

The last stage of the procedure was to re-record the spliced master tapes into the four experimental tapes as specified in the design. Once again the transfer was accomplished from the Ampex

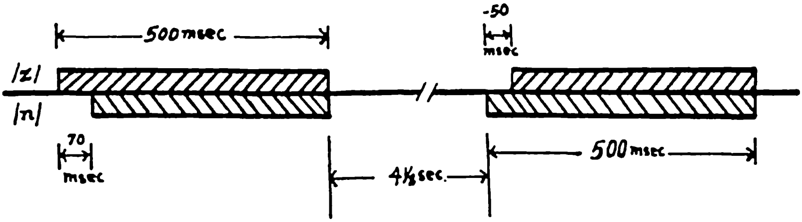


Figure 5. Schematic diagram of a typical stimuli sequence showing the stimulus pair, z/n, with onset-differentials of 70 and −50 msec respectively. (This figure does not depict a tape segment.)

300 to the Ampex 601 after providing for constant VU-level recording. A special brand of low-print magnetic tape[4] was used for all master recordings. Finally, the experimenter dubbed in the word "signal" 8 seconds before the onset of each new stimulus pair.

TESTING OF SUBJECTS

Subjects for the experiment were 28 females and 20 males ranging in age from 16 to 28 years. The median age for the group was 19 years. Prospective subjects were drawn from two sources, a group of high school senior boys attending a one-week institute at the Purdue Speech and Hearing Clinic, and students newly-enrolled in an undergraduate voice and articulation course at Purdue. All subjects indicated that they were free from known history of hearing loss or pathology. Most of the subjects were native to the Midwestern United States and all were speakers of the English language since childhood. Prospective subjects were screened for possible hearing losses by pure-tone air-conduction audiometry. The criterion for rejection was inability to pass a right-ear audiometric sweep check at 15 db for any one of the following frequencies: 250, 500, 1000, 1500, 2000, 4000, and 6000 cps. A Maico Model F1 portable audiometer was used for the testing.

[4] Scotch Brand, Type 131-12, Minnesota Mining and Manufacturing Company, Saint Paul, Minnesota.

Prospective subjects were scheduled in groups of four, although eliminations and scheduling conflicts necessitated testing smaller groups on some occasions. The experiment was conducted in the sound-treated auditory training room of the Purdue Clinic. A curved, hard-top counter equipped with several Grayson-Stadler CB 6012 volume-control boxes was used for the testing procedure. Cardboard partitions were constructed making four individual writing compartments. The staggered seating arrangement and the general testing environment may be seen in Figure 6. When unoccupied the writing compartments had the following sound-level readings ("A" and "C" weighting networks, respectively): Compartment #1, 38 and 52 db; Compartment #2, 40 and 54 db; Compartment #3, 40 and 54 db; and Compartment #4, 42.5 and 56 db. Minor differences were attributed to the relative position of the compartments in relation to the playback recorder which produced some operating noise. The level of recorded stimuli and the effectiveness of seal of the phone sockets were such that this ambient noise level was considered satisfactory for the experiment. Readings were obtained from a General Radio 1551-B Sound Level Meter.

Each compartment was equipped with a set of TDH-39 headphones fitted with Grayson-Stadler semi-plastic earphone sockets. Each headphone set was connected to a near-by Grayson-Stadler CB 6012 control box, each of which, in turn, was connected to the output of the Ampex 601 tape recorder located on an adjoining counter-top to the rear of the testing compartments. Only the right phone of each headset was active. Each phone was calibrated to deliver the tape-recorded signal at 75 db SPL (re .0002 microbar). The calibration procedure did not entirely eliminate small differences in output among the four receivers. However, since all positions were used an equal number of times for a given sequence of stimuli, no known bias was introduced by the headphones.

The first stage of the testing procedure was to screen the hearing and complete the individual data requirements for each prospective subject. Persons not qualifying were excused. Each subject was

Figure 6. Photograph of the seating and testing arrangements. The playback recorder may be seen at center left.

given his response sheet with the appropriate pre-marked experimental sound pairs appearing at the top of each double column. The oral instructions to the group indicated the four stimulus pairs of the test, the decision-making task involved, and the general testing procedure. Subjects were told to write or print in the space provided the letter symbolizing the first sound heard. It was emphasized that they must write a choice after each presentation, resorting to a "best guess" if necessary. In the general instructions and before each new stimulus-pair presentation, subjects were cautioned to begin on time, keep up, and to write after each presentation their best judgment of the first sound heard.

After procedural questions were answered, the subjects were seated comfortably at the counter and fitted with the headsets. The experimenter, also equipped with a headset, then engaged the Ampex 601 at the proper pre-marked point on the experimental tape after an indication that the subjects were ready. The tape was begun about 2 seconds before the word "signal", which preceded the first condition by 8 seconds. After completion of a 40-item stimulus pair sequence (about 3 minutes) subjects were given a 3-minute rest, timed with a Kodak timer. Before listening resumed, the experimenter informed the subjects of the next stimulus pair involved and repeated the condensed instructions noted earlier. Whenever a shortened form of the sound symbol was permitted, for example "T" for "TH" or "H" for "High", the subjects were so informed. Before leaving, the subjects were asked not to disclose the purpose or procedure of the experiment to other known prospective participants. Total testing time was about 35 minutes.

The main experimental data collected (excluding the 80-msec practice judgments and the zero-onset-disparity conditions) consisted of 32 responses for each of four stimulus pairs for each subject. A phoneme pair was judged by 12 subjects, providing 384 judgments per pair or 48 judgments per temporal condition. The pure-tone pair was judged by all 48 subjects in order to obtain a constant-performance criterion. This condition, therefore, yielded a total of 1536 judgments, or 192 judgments per temporal condition.

RESULTS

Several statistical procedures for treating the data were considered, but in most it proved to be impossible to meet the basic assumptions underlying the tests. Even non-parametric procedures were demonstrated to be unsuitable, because of an excessive number of zero-frequency data cells. It was therefore concluded that the experimental results could be handled most satisfactorily with descriptive techniques, a conclusion evidently reached by other investigators of these phenomena as well.

The correct responses obtained by each subject for each experimental condition were tabulated. These tabulated data were then analyzed both for individual scores of the four repetitions of each temporal condition and for the pooled scores of 48 responses per temporal condition for a 12-subject group. Because the former measure, individual results on the four repetitions, offered independent performance criteria, this procedure was considered the preferred method. The median was selected as the most suitable measure of central tendency inasmuch as the distributions tended to be skewed in the direction of 100%.

Table II is a summary of the performance on the phoneme stimulus pairs of all subjects under all conditions of onset disparity. Both the total correct scores as reflected in the 48 judgments by each subject group and the medians (percentages) of the 12 subjects within a group are indicated. It should be noted that each instance in which the first member of the phoneme pair preceded the second is indicated as positive onset disparity. The columns on the negative side, conversely, reveal the results when the second phoneme member preceded.

A median score of 75% is considered the criterion level for

TABLE II

Summary of total correct responses (of 48) and median scores (of 12 subjects) for each phoneme pair according to the experimental onset disparity of the speech sounds

Phoneme Pair		Onset Disparity (msec)							
		−70	−50	−30	−10	+10	+30	+50	+70
m/r	Total	44	47	40	42	41	46	44	46
	Mdn. (%)	100	100	100	100	100	100	100	100
n/r	Total	43	41	41	44	36	46	44	44
	Mdn. (%)	100	100	100	100	75	100	100	100
v/l	Total	36	29	37	31	30	42	38	45
	Mdn. (%)	75	75	75	75	50	100	100	100
v/r	Total	36	32	38	33	26	39	38	41
	Mdn. (%)	75	75	87.5	75	50	87.5	87.5	100
z/r	Total	32	30	21	29	30	35	37	35
	Mdn. (%)	75	75	37.5	50	50	75	87.5	75
ð/r	Total	37	33	29	25	32	35	30	38
	Mdn. (%)	75	50	75	62.5	50	87.5	75	100
n/ð	Total	37	39	41	44	13	30	40	42
	Mdn. (%)	87.5	75	87.5	100	25	62.5	87.5	100
z/n	Total	46	40	31	23	25	22	30	29
	Mdn. (%)	100	100	62.5	50	50	37.5	62.5	75
l/z	Total	29	23	28	32	25	27	36	40
	Mdn. (%)	62.5	50	50	50	50	50	75	87.5
l/ð	Total	27	26	24	20	26	32	31	32
	Mdn. (%)	50	37.5	50	37.5	50	75	62.5	75
l/r	Total	22	27	28	33	20	24	28	37
	Mdn. (%)	50	62.5	50	75	50	50	50	87.5
m/n	Total	27	21	21	24	27	29	30	27
	Mdn. (%)	50	37.5	37.5	37.5	50	62.5	50	62.5

correct performance for these data.[1] The performances between the chance level of 50% and the criterion score have been termed by Hirsh (40, p. 14) as composing "the interval of uncertainty". The arrangement of the data in Table II is generally in rank order from "superior" at the top to "poor" at the bottom as determined by 75% or better median scores.

Table III summarizes the results obtained from the 0-msec condition, simultaneous onset, of each stimulus pair. In addition to the division of choices between the members of a stimulus pair, whole-number proportions of these responses are indicated. These data, it will be seen, have proven valuable as an aid to evaluating performances, insofar as identification of a dominant member of a pair is indicated.

TABLE III

Summary of the distribution of choices and their relative proportions for the individual members of a stimulus pair at the 0-msec (simultaneous) condition. The pure tones, PTs, are shown by subject groups

Stimulus Pair	*Choice*	*Proportion*	*Stimulus Pair*	*Choice*	*Proportion*
m/r	7/41	15/85	l/z	22/26	46/54
n/r	6/42	13/87	l/ð	31/17	65/35
v/l	18/30	38/62	l/r	16/32	33/67
v/r	18/30	38/62	m/n	26/22	54/46
z/r	20/28	42/58	PTs (1)	22/26	46/54
ð/r	26/22	54/46	PTs (2)	29/19	60/40
n/ð	9/39	19/81	PTs (3)	23/25	48/52
z/n	25/23	52/48	PTs (4)	23/25	48/52

Figures 7 and 8 present in pie-graph form the median performance scores of the 12 subjects who judged the m/r, Figure 7, and n/r, Figure 8, phoneme pairs. The range of individual scores

[1] Among the investigators who have employed the 75% criterion are Kinney (48), Chistovich (as reported by Broadbent and Ladefoged [11]), Hirsh (41), Hirsh and Sherrick (42), Tobias and Zerlin (89), and Efrom (20).

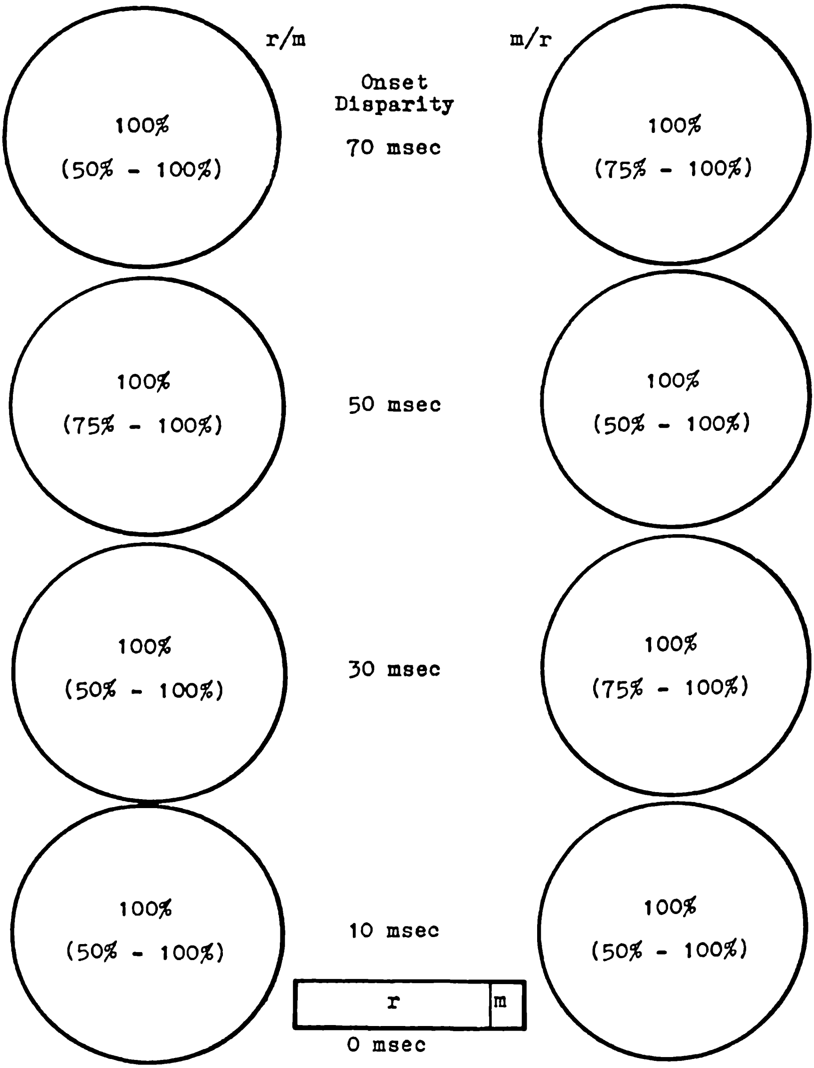


Figure 7. Median scores and ranges (in parentheses) for the m/r phoneme pair by temporal condition. Simultaneous onset results are shown proportionally by phoneme in the 0-msec box.

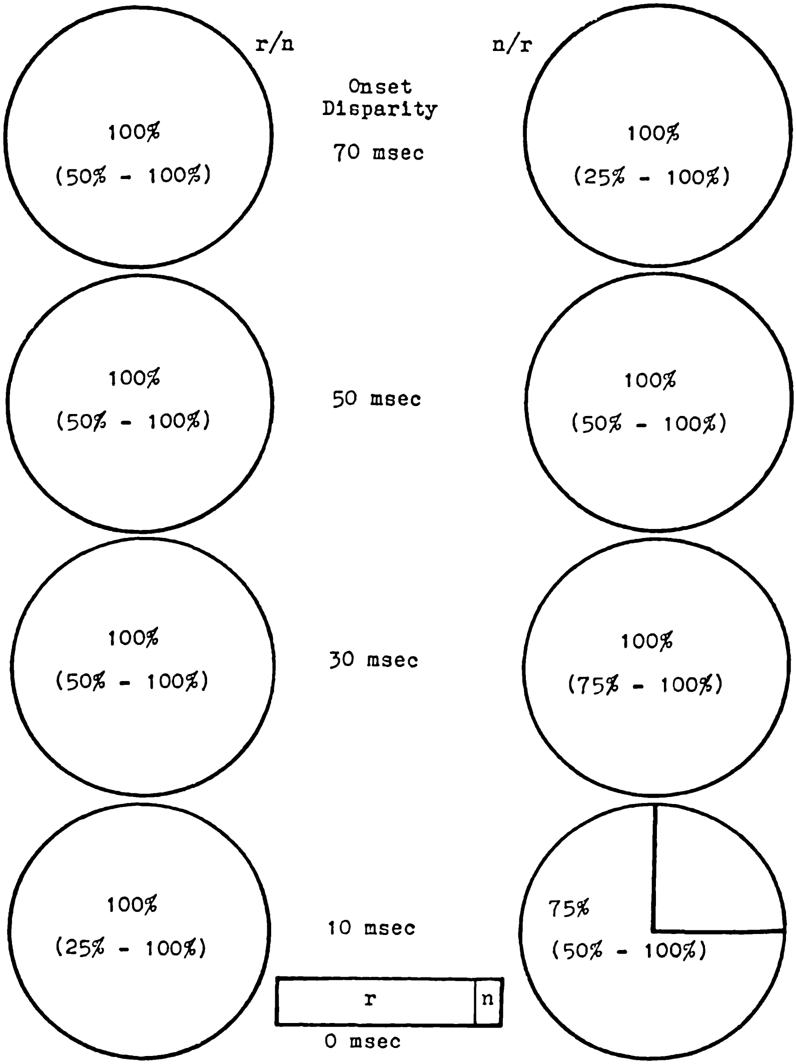


Figure 8. Median scores and ranges (in parentheses) for the n/r phoneme pair by temporal condition. Simultaneous onset results are shown proportionally by phoneme in the 0-msec box.

for each temporal condition is shown in parentheses beneath each median score. Fine discrimination by the subjects at all temporal conditions is evident in both Figures 7 and 8. The only exception to the 100% median score was the 75% performance for the n/r sequence at the shortest onset separation, 10 msec. In addition to the universally high medians, it can be seen that the range of scores is limited largely to 50% or better. With reference to the 0-msec box, simultaneous-onset selections, it should be observed that both figures show similar proportions, and in both the /r/ is the strong choice.

Figures 9 and 10 show the great similarity of results obtained for the stimulus pairs v/r and v/l. This similarity between the pairs is evident in (a) the criterion (75%) performance in each instance for all conditions except the 10-msec v/semi-vowel (both /r/ and /l/) which was chance level for each, (b) identical 0-msec proportional choices for the semi-vowel in relation to the /v/, (c) the rather close correspondence of ranges of scores by condition, and (d) the supra-criterion medians achieved for the fricative-first onset sequences (30, 50, and 70 msec) compared to the performance at criterion level for all /r/-first sequences except at 30-msec r/v.

The results of the z/r and ð/r pairs, Figures 11 and 12, though similar to the preceding figure in which /r/ was also paired with a fricative, show more evidence of the effects of short inter-onset intervals. For these pairs criterion level was obtained in both instances at 30 msec in the fricative-first condition. With the /r/ preceding, however, results varied between the pairs with criterion level obtained at 50 msec for r/z and at the 70- and 30-msec (but not 50-msec) conditions of ð/r. The ranges of scores for both pairs were varied and quite wide.

Figure 13, n/ð, demonstrates a decided difference in performance according to the phoneme initially presented. The n/ð sequence was a steadily increasing function with the increase in onset disparity achieving criterion level at 50 msec. On the other hand, results were at or above criterion for all conditions of the ð/n sequence with 10-msec performance superior to that at 70-msec.

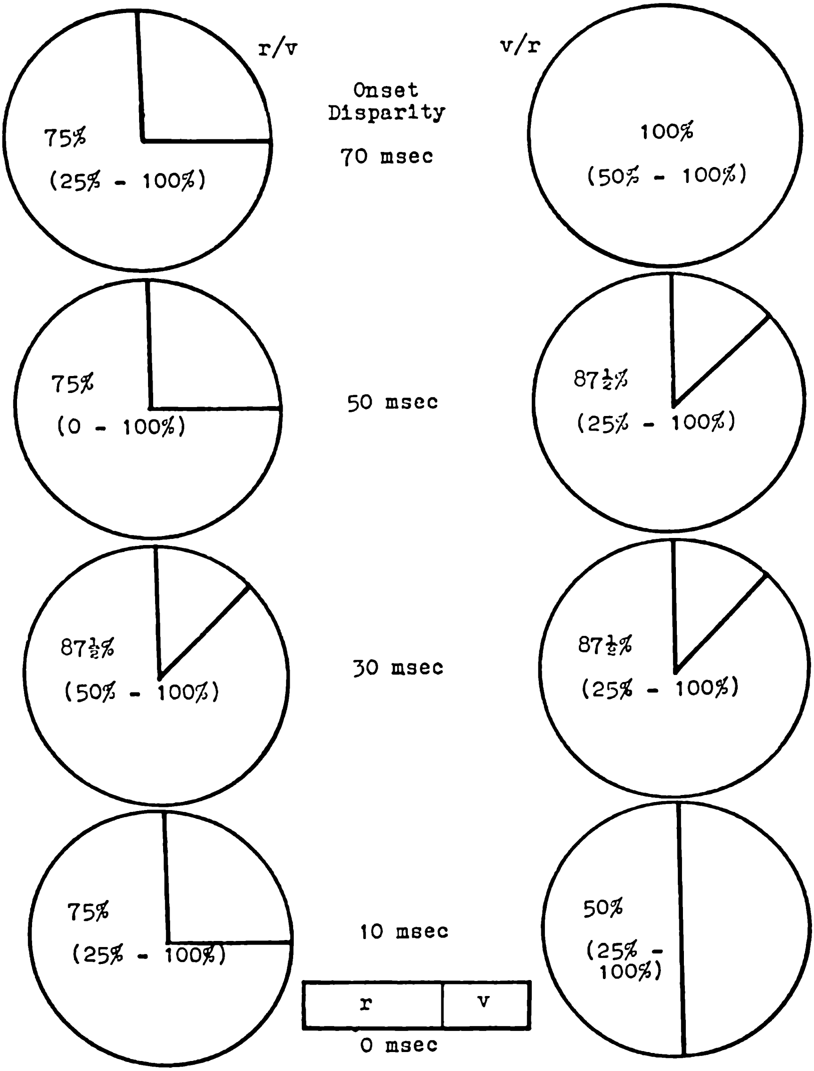


Figure 9. Median scores and ranges (in parentheses) for the v/r phoneme pair by temporal condition. Simultaneous onset results are shown proportionally by phoneme in the 0-msec box.

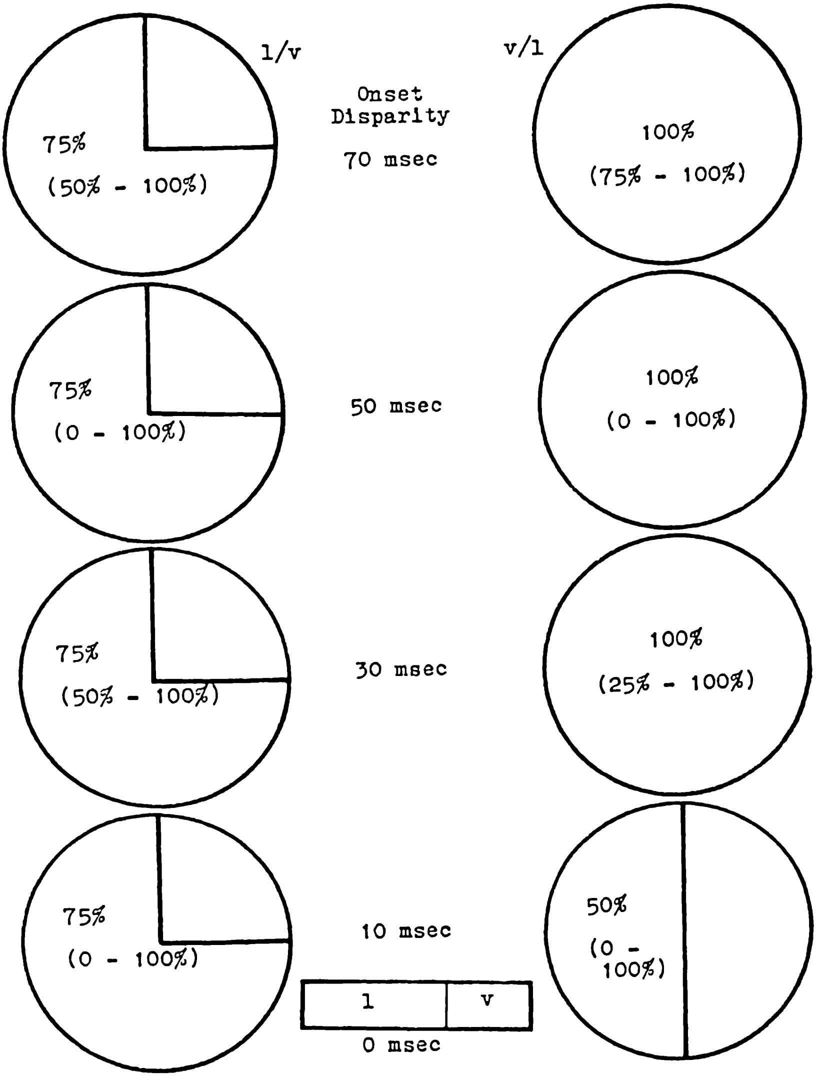


Figure 10. Median scores and ranges (in parentheses) for the v/l phoneme pair by temporal condition. Simultaneous onset results are shown proportionally by phoneme in the 0-msec box.

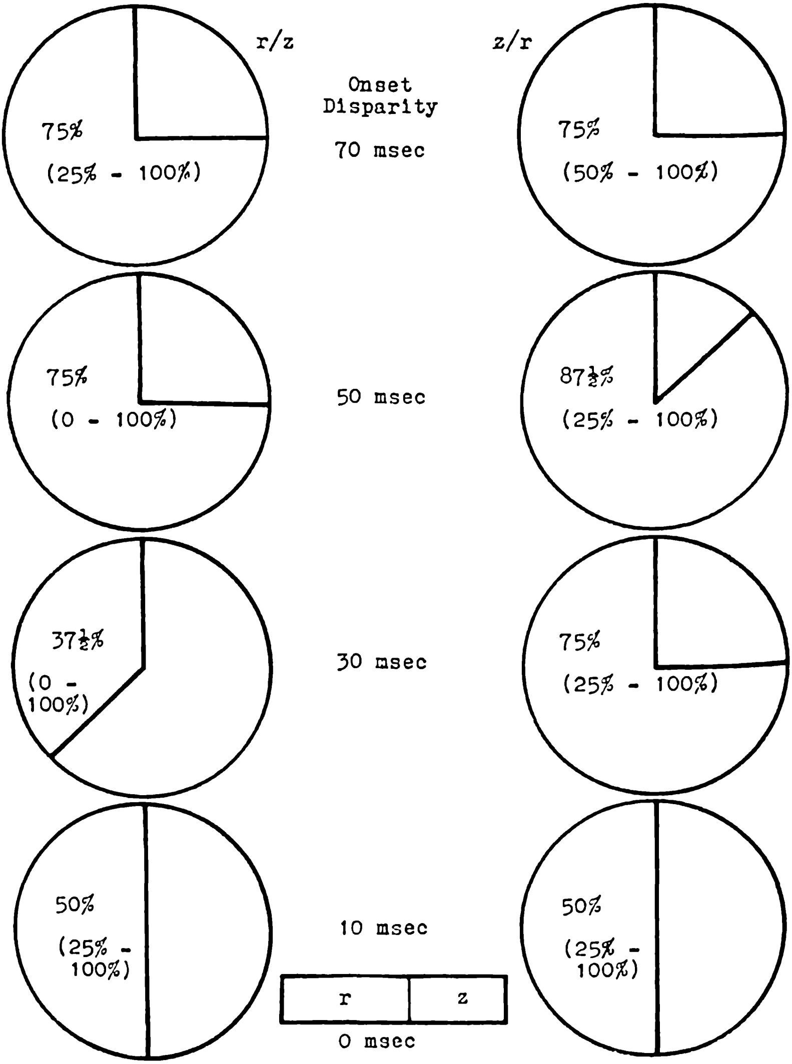


Figure 11. Median scores and ranges (in parentheses) for the z/r phoneme pair by temporal condition. Simultaneous onset results are shown proportionally by phoneme in the 0-msec box.

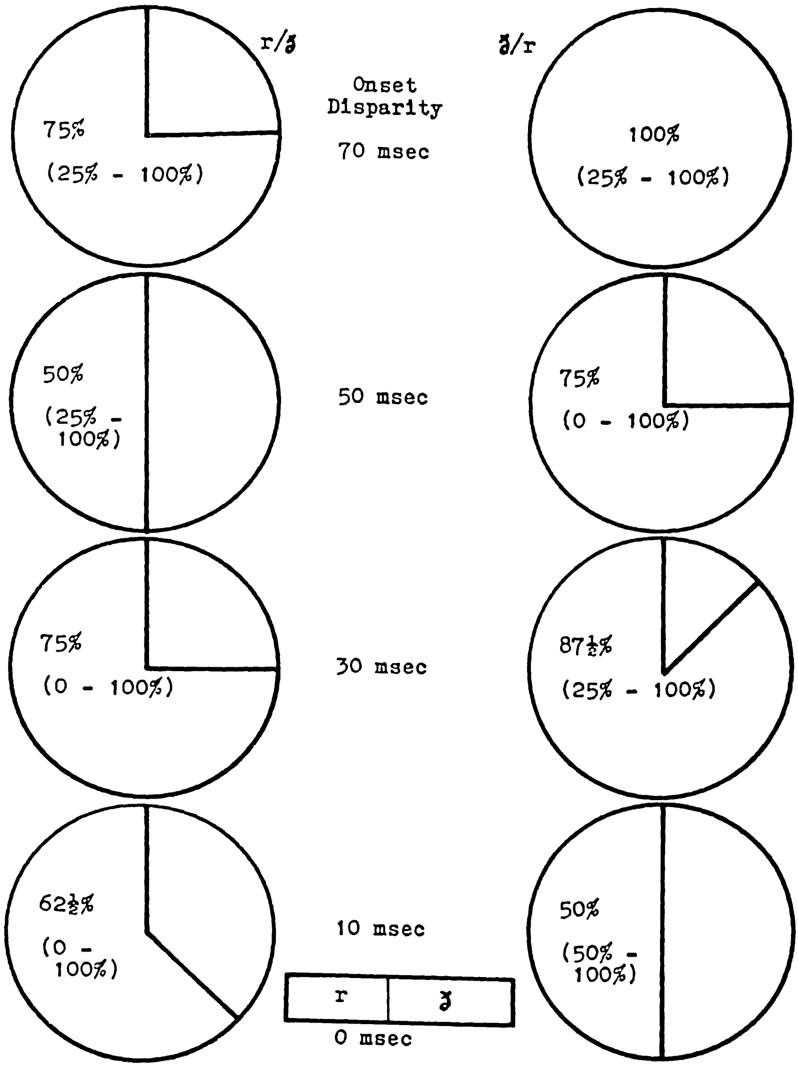


Figure 12. Median scores and ranges (in parentheses) for the ð/r phoneme pair by temporal condition. Simultaneous onset results are shown proportionally by phoneme in the 0-msec box.

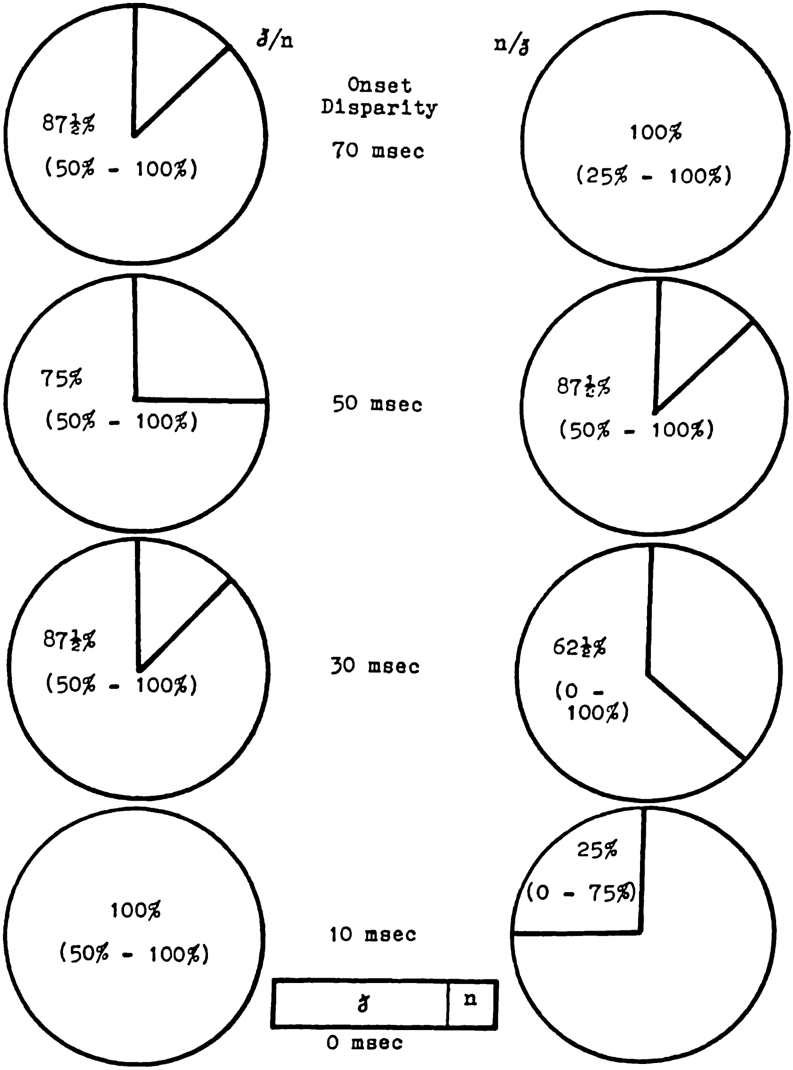


Figure 13. Median scores and ranges (in parentheses) for the n/ð phoneme pair by temporal condition. Simultaneous onset results are shown proportionally by phoneme in the 0-msec box.

Once again, it will be noted in Figure 13 that the 0-msec box reveals a disproportionate selection for /ð/ as opposed to /n/.

The results for z/n, Figure 14, were somewhat different from a similar pairing of fricative/nasal as seen in Figure 13. The influence of the temporal separation of onsets seems more apparent in Figure 14 than in any previous graphical figure. In both onset positions median scores reflected a slow rise to criterion level, with n/z crossing the 75% point at 50 msec and z/n doing so at 70 msec. The score ranges also reflected the somewhat earlier resolution of the n/z sequence.

With Figures 15 and 16 the first examples may be seen of results failing to reach criterion at maximum onset disparity. In these figures, l/ð and l/z respectively, the fricative presented initially resulted in sub-criterion performances. With the /l/ leading, a 50-msec separation was needed to achieve criterion in Figure 16 whereas in Figure 15 this level was obtained at 30 msec, lost at 50 msec, and regained at 70 msec.

Figure 17 shows the generally poor results of pairing two semivowels, /r/ and /l/. The 10-msec r/l not only shows the highest median for the sequence, but also the highest range. With the l/r sequence 70 msec was required to resolve the order correctly.

The final phoneme pair, m/n, is presented in Figure 18. Performance at and around the chance level is apparent at all temporal conditions of both onset sequences.

Each subject group was tested with the pure-tone pair, 1200/250 cps, as the final stimulus presentation. The results on the pure tones by subject group are shown in Table IV and Figure 19. The table gives the median scores as well as the total correct responses of each group and of the entire 48 subjects. Figure 19 shows the modes of the median scores obtained for each of the four subject groups. Each mode defines two, three, or four medians as a measure of central tendency. The actual number of medians represented is indicated before the mode, and the remaining one or two medians are shown in parentheses beneath. The simultaneous-onset results are shown as four 0-msec boxes representing proportional choices by each subject group (see Table III).

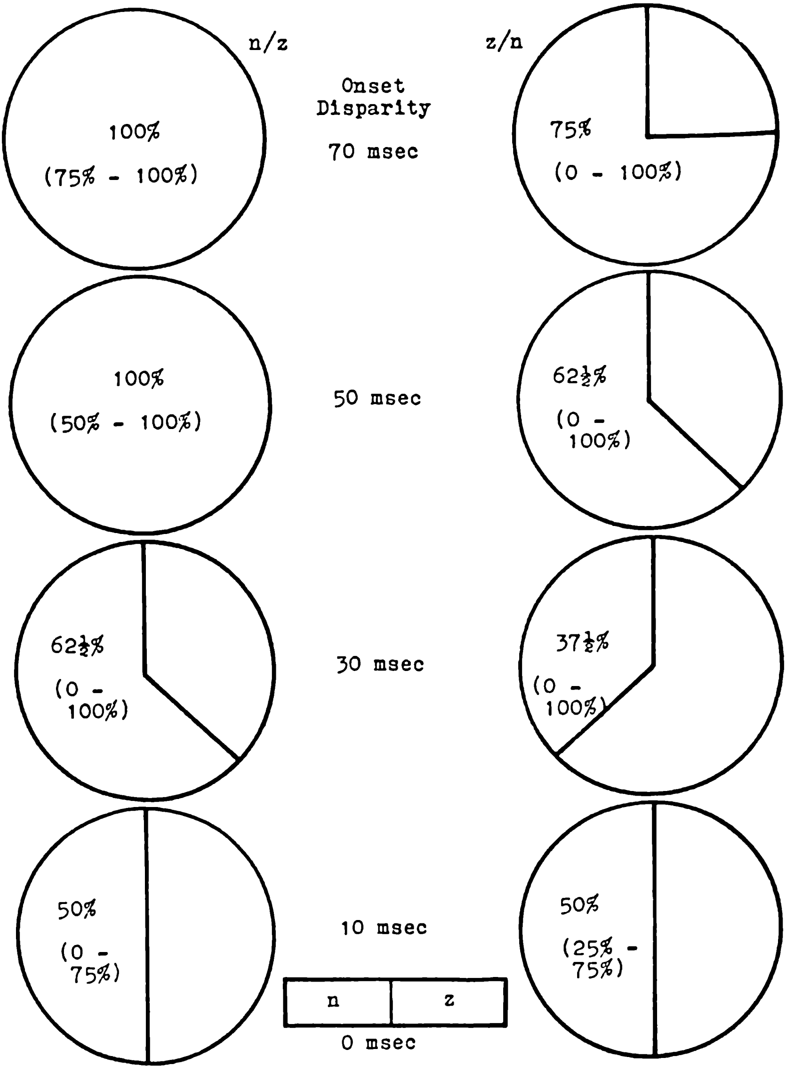


Figure 14. Median scores and ranges (in parentheses) for the z/n phoneme pair by temporal condition. Simultaneous onset results are shown proportionally by phoneme in the 0-msec box.

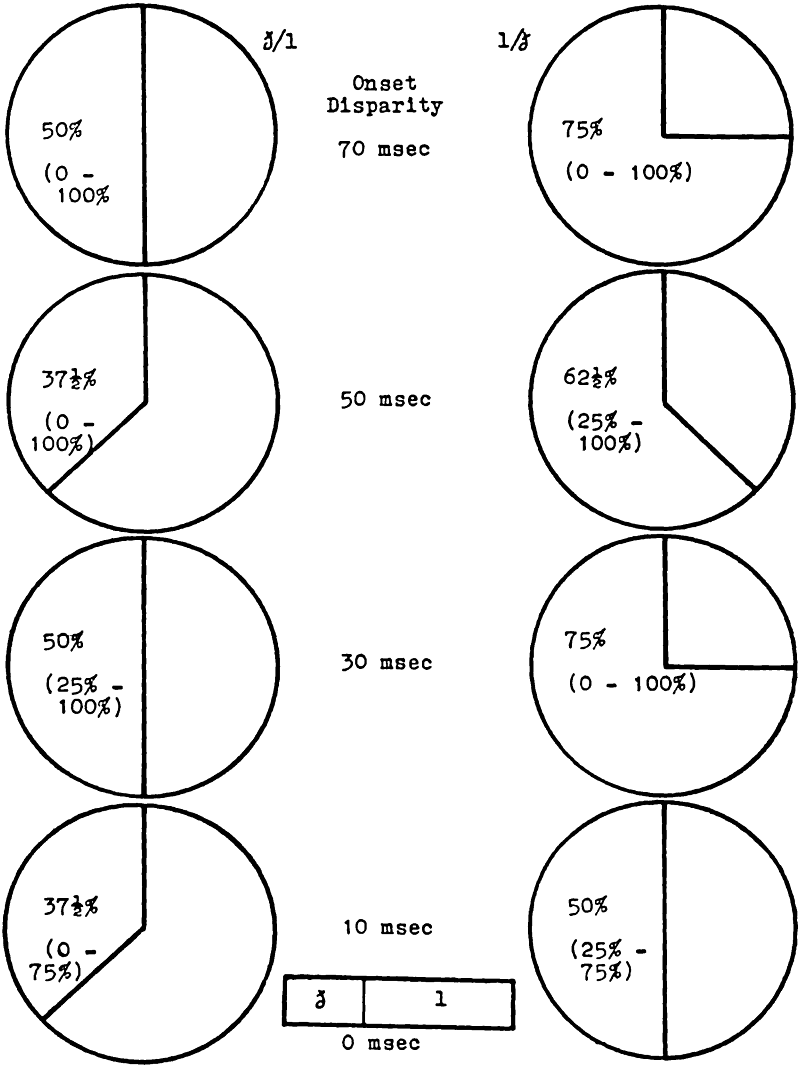


Figure 15. Median scores and ranges (in parentheses) for the l/ð phoneme pair by temporal condition. Simultaneous onset results are shown proportionally by phoneme in the 0-msec box.

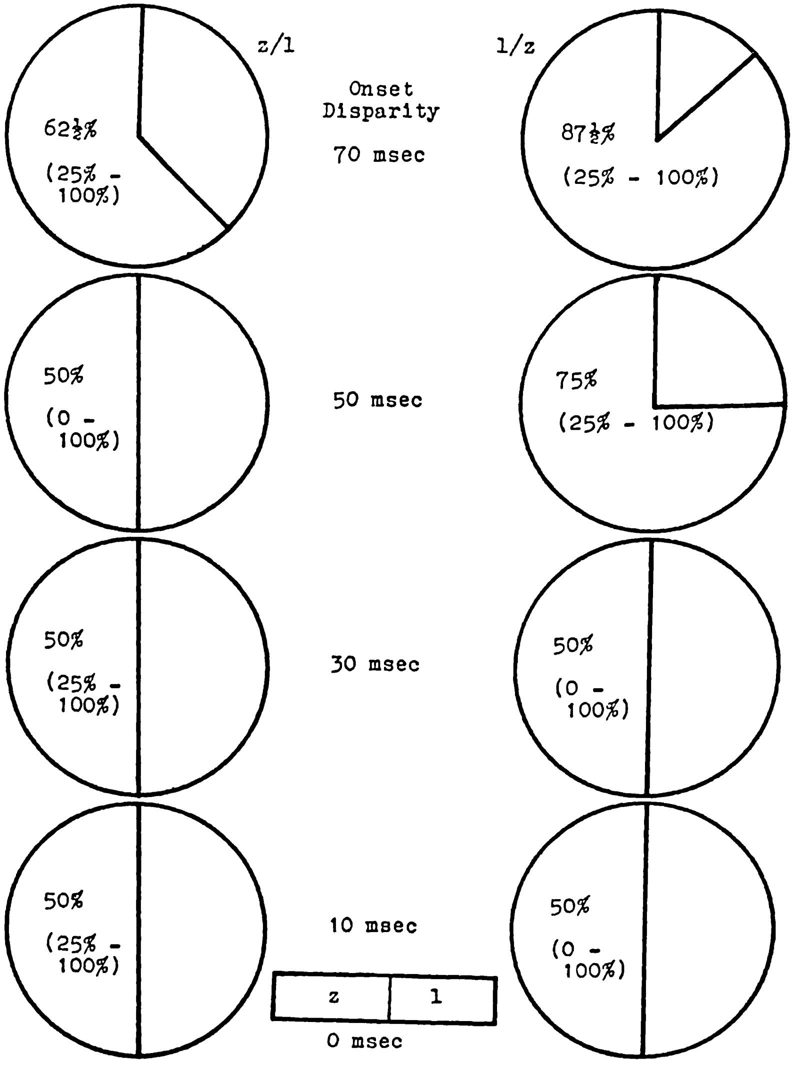


Figure 16. Median scores and ranges (in parentheses) for the l/z phoneme pair by temporal condition. Simultaneous onset results are shown proportionally by phoneme in the 0-msec box.

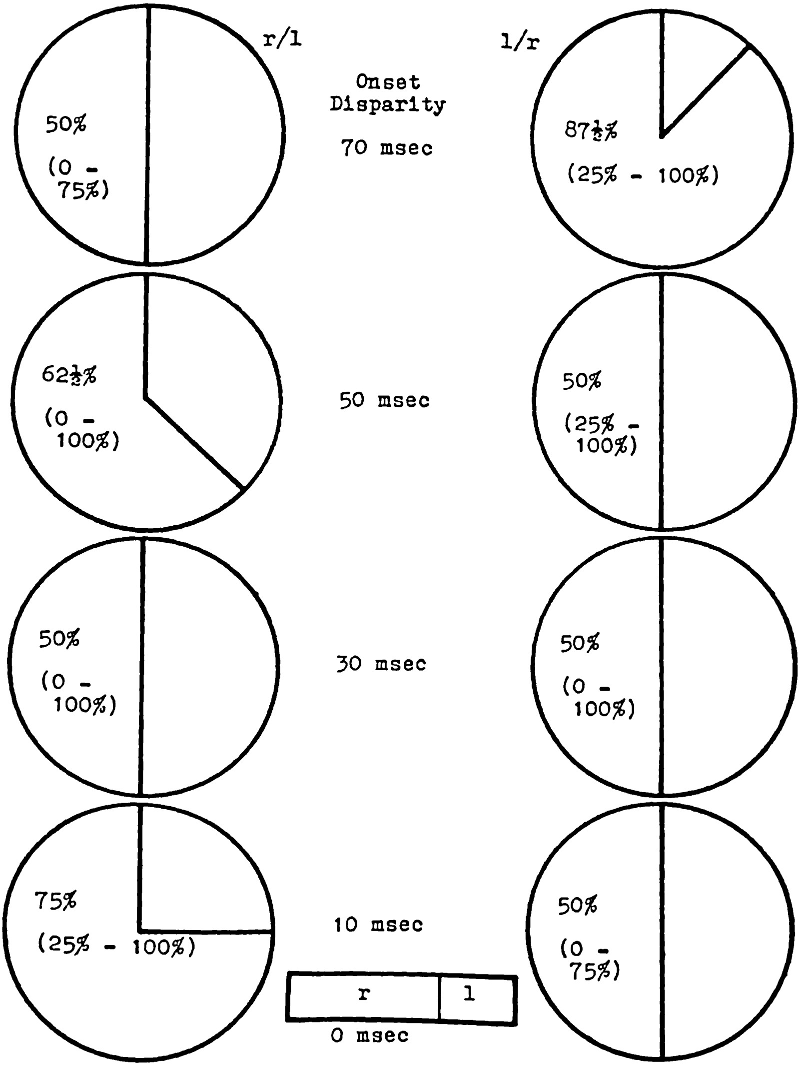


Figure 17. Median scores and ranges (in parentheses) for the l/r phoneme pair by temporal condition. Simultaneous onset results are shown proportionally by phoneme in the 0-msec box.

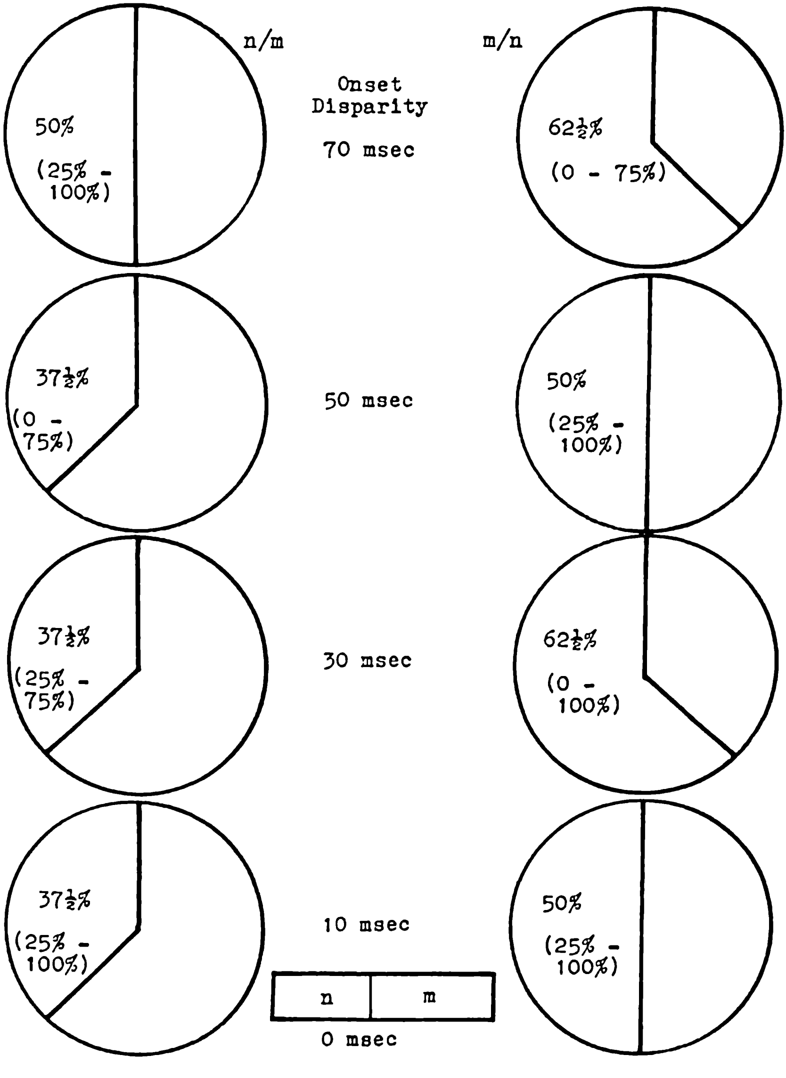


Figure 18. Median scores and ranges (in parentheses) for the m/n phoneme pair by temporal condition. Simultaneous onset results are shown proportionally by phoneme in the 0-msec box.

TABLE IV

Summary of total correct responses (of 48) and median scores for the pure-tone pair for each of four groups of 12 subjects. Also shown are the total correct responses of all 48 subjects

Group		Onset Disparity (msec)							
		-70	-50	-30	-10	+10	+30	+50	+70
1-12	Total	34	32	22	21	25	27	33	34
	Mdn. (%)	75	75	50	25	62.5	50	75	75
13-24	Total	30	21	19	22	23	27	30	35
	Mdn. (%)	75	50	37.5	50	37.5	50	50	75
25-36	Total	39	30	33	28	19	20	28	31
	Mdn. (%)	75	62.5	75	62.5	37.5	37.5	50	75
36-48	Total	29	35	27	28	23	30	27	32
	Mdn. (%)	50	75	50	50	50	62.5	50	75
All-Subject	Total	132	118	101	99	90	104	118	132

In the sub-criterion or "uncertain" intervals, the performances of the groups may be seen to vary somewhat. At the 75% point, however, there is considerable unity in the results for all four subject groups. Performance below 50 msec is best described as chance, whereas the 50-msec results show near-criterion discrimination, and at 70 msec criterion level is almost completely achieved. It may be noted that 100% medians are missing from the pure-tone results. Assuming that Figure 19 does represent the central tendency of performance for the 48 subjects, a comparison with the 12-subject phoneme results would reveal a similarity with Figures 14, z/n, or 16, l/z. In terms of rank, temporal resolving performance on the pure tones was inferior to that on half of the phoneme pairs involved in the investigation.

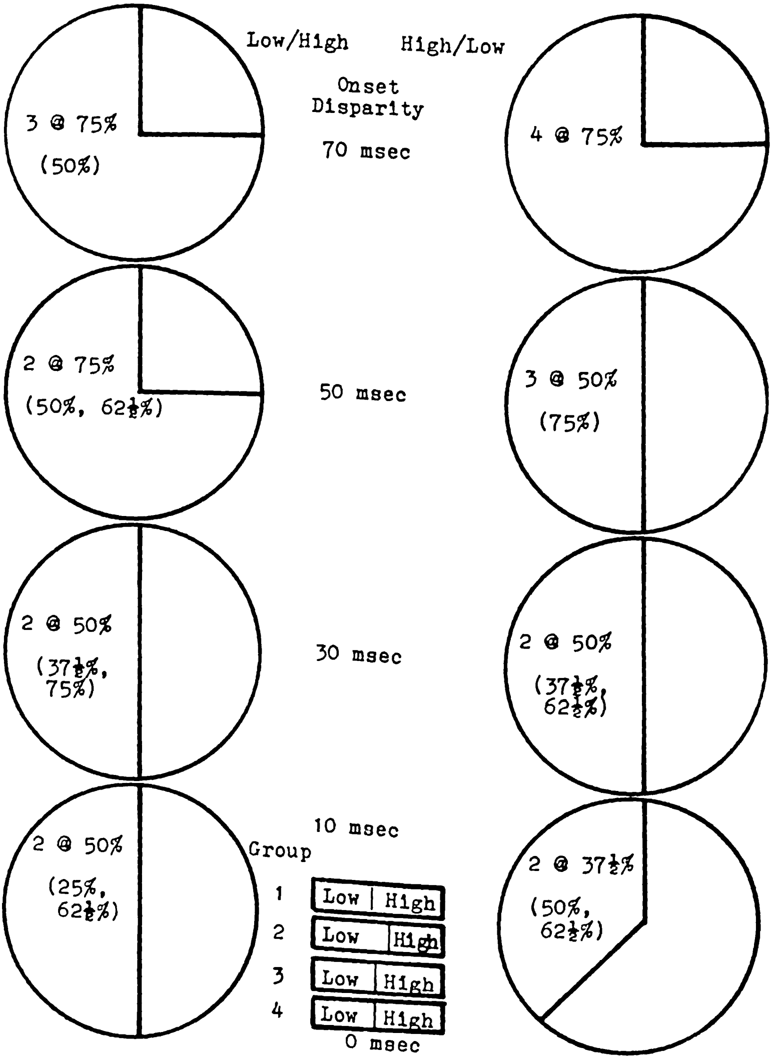


Figure 19. Modes of the median scores within four subject groups on the pure-tone pair. The number of medians comprising the mode is indicated and non-modal medians are shown in parentheses. The 0-msec proportions are shown by group.

Data from the initially presented four repetitions of each stimulus pair with 80-msec onset separation of the component sounds were not considered in the experimental results. It is interesting to note, however, that median scores were 100% for seven phoneme pairs and three of the four pure-tone groups. The remaining six median scores were 75%.

DISCUSSION

In analyzing and evaluating the results of this investigation a functional approach to speech perception is employed. Therefore, rather than limit analysis to an acoustic or an informational or a linguistic frame of reference, the investigator will view any single result from the frame of reference that appears to offer the most logical explanations.

THE INFERIOR PAIRED NASALS

Without doubt the easiest results to elucidate are those of the chance level m/n pair, Figure 18. Such a pair when found contiguous in English is either evidenced in division at syllable boundaries, *gymnasium* and *inmate*, or is sounded as a single nasal, *column*.

A common metathesis example, *alúnimum* or *alúnimun*, might thus be explained in the poor temporal resolution results for this pair. It is interesting to note that the British say *alúmińium*, a word which no doubt creates less pronunciation problems. The easier pronunciation might be explained on the basis of the strong accent on [mɪn], which serves to increase temporal distance between /m/ and /n/.

From the acoustic frame of reference there is also evidence to explain the poor m/n results. Malécot (61), using a technique which involved separating and recombining taped segments of utterances and then presenting the new stimuli to phonetically-naive subjects for identification, found that the *place* of articulation of /m/, /n/, and /ŋ/ is conveyed principally by the transitions of

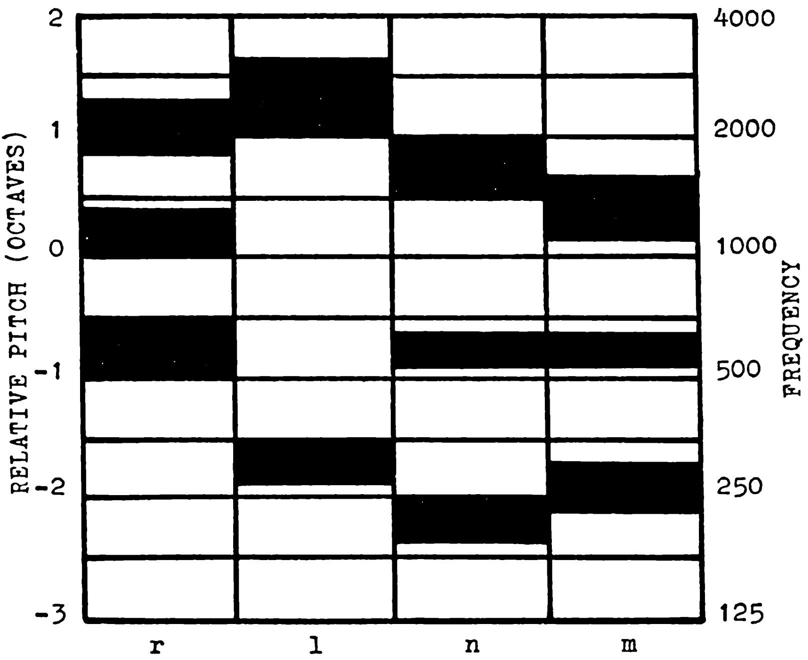


Figure 20. Characteristic resonance positions for spoken semi-vowels and nasals. (After Fletcher 27, p. 63.)

the adjoining vowel formants. The vowel transition was found to be decisive for the perception of the combination nasal-plus-vowel (/m/ plus /æ/, separating the vowel from the combination /næ/), whereas the perception of final nasals was not to the same extent dominated by the vowel transition. Liberman (51), in his work with synthetic speech, observed that the nasal cue from the fixed resonator in the nose does not provide much basis for distinguishing the sounds within the class of nasal consonants. He suggested that to accomplish such a task the listener must rely on the transitions of the second and third formants. Such transitions were missing from the isolated utterances of the present investigation.

Finally, attention is invited to Figure 20, an adaptation of an analysis of formant frequencies reported by Fletcher (27, p. 63) from his own data and those of others. Fletcher emphasizes, however, that there are considerable variations from these values by different speakers and also by the same speaker at different times. In this figure it may be seen that the similarity of the resonant frequencies of /m/ and /n/ lends additional support for the obvious discrimination difficulty reflected by the results.

THE SUPERIOR R/NASAL PAIRS

A most significant finding of the experiment was the superior performance indicated when the /m/ and /n/ were paired with /r/, Figures 7 and 8. It is believed that two influences functioned to produce the fine discrimination results apparent in these figures. One of these influences, superior nasal/r discrimination, will be discussed in the succeeding section. The other, herein termed "frequency masking", is considered in the present section because it is believed to explain in part the results of the r/nasal pairs as seen on the left-hand portions of Figures 7 and 8.

The design employed for this investigation was one of overlapping phoneme sequence. Thus, the onset of the second phoneme in each presentation of the experiment was also the onset of concurrent presentation of the two stimuli, see Figure 5. Such a design leads to the possibility of introduction of masking effects. Peterson (73) has observed that the amount of interference is variable:

> Two physical stimuli or signals which are presented simultaneously are never indistinguishable in all respects, since they cannot separately occupy the same space at the same time. If perceived at all, they can be identified as separate. It is certain aspects, qualities, or properties of these stimuli or signals which cannot be distinguished.

It is suggested that because of the relative dominance of a certain phoneme in relation to the phoneme with which it was paired, "certain aspects, qualities, or properties" of the less dominant member became indistinguishable. Since the stimuli were all recorded at a constant record-level, it is assumed that the dominating characteristic must be one of the component frequencies of the sounds. Hirsh (41) observed some masking effects with certain paired sounds of close frequencies.

With regard to the r/nasal sequences, it is believed that the /r/ was the dominant member of the phoneme pair. Observations on the part of several listeners and the experimenter indicated that when the /r/ led either nasal, the impression was predominantly one of /r/ alone. Further evidence for the relative dominance of /r/ may be seen in the disproportionately large selection of the sound for the simultaneous onset condition as seen in Table III and the 0-msec boxes of Figures 7 and 8. Finally, reference to Figure 20 lends support to the potential for dominance of the /r/ preceding and subsequently overlapping /m/ or /n/. In this figure it may be seen that the lower resonance position (first formant) of the /r/ could effectively mask the narrower second formant band of either of the two nasals. In addition, the high-frequency second and third formants of the /r/ might interfere with perception of the third formants of the nasals.

Now assuming that in the forced-choice experimental session involving temporal resolution of the m/r and r/m sequences, the listener could (a) accurately identify the /m/-first sequence and (b) hear only a dominant /r/ utterance in presentations when the /m/ did *not* precede, results for the r/m sequences would be highly accurate. It is believed that this combination of circumstances existed for both the r/m and r/n sequences, and that the excellent results shown on the left side of Figures 7 and 8 do not reflect accurate resolution, but rather the simple alternative choice to a sequence which *did* reflect accurate resolution. The actual performance on the r/m and r/n sequences, therefore, is a matter of some speculation.

It might be pointed out here that the apparent existence of a masking effect with some phonemes in the laboratory does not, it is believed, preclude the importance of this effect as a determinant of phoneme duration or sequence in non-experimental speech perception. A phoneme possesses certain physical characteristics which are no doubt generally in evidence in the speech samples herein employed as well as in normal oral language usage. Therefore, if experimentally-imposed temporal restrictions produce a masking effect, this same effect is likely to be in evidence in the perception of certain linguistic sequences. It is suggested that just as given articulatory sequences are contrary to facilitated motor activity, so too might certain auditory sequences be contrary to perceptual ease. Frequency masking in a sequential relationship may therefore affect perception of the linguistic phoneme complex in very important ways.

THE SUPERIOR NASAL/R PAIRS

In contrast to the r/nasal sequences, the nasal/r sequences did seem to be resolved with very high accuracy when onsets were separated by as little as 10 msec, Figures 7 and 8. When one considers that awareness of successiveness of clicks and other sounds requires from 2 to 10 msec (46, p. 613; 92; 74, p. 302; 34; 77), then these results with phonemes are particularly impressive since the minimum resolution time was not established. Just how much below 10 msec resolution would continue to be accurate is a question unanswered.

As discussed in the previous section, there is evidence that the r/m and r/n results showed masking on the part of the initial /r/. Such a result would not preclude an effect on the reverse, initial /n/, sequence except to suggest that from the point of onset of /r/ the concurrent sound complex was heard as a dominant /r/ with little if any perception of /n/. Such a relationship would be, therefore, similar to a 10-msec segment of /n/ or /m/ followed by a 480-msec segment of /r/ without apparent overlapping.

As shown in these figures, it seems reasonable to believe that the m/r and n/r tests were virtually duplicates. The inability of subjects to discriminate within class nasals lacking vowel transients was pointed out in the section dealing with the paired nasals. Therefore, the nasal could be equally well heard as either /m/ or /n/ depending, no doubt, on the experimenter's requested identification. Malécot (61) observed that exclusively steady-state resonances of nasals paired with steady-state vowels showed a clear loss of identifiability as compared with unaltered transitions. He attributed the loss to absence of transitions. Thus, it would appear that explanations for nasal/r are equally suited to either n/r or m/r since the stimulus pairs were virtually identical.

The important question raised by the results of the nasal/r sequences is why or how the nasal was perceived and properly ordered in relation to the /r/? The premise to be explored here is that the nasal and /r/ were not two independently perceived and ordered phonemes, but rather they were a single perceptual unit or temporal package.

The assumption that /mr/ and /nr/ form identifiable units and are thereby more readily perceived might be explained by the suggestion that the two-sound cluster is acoustically more acceptable and hence more expeditiously handled by the central ordering processes. However, all of the phonemes in the study were consonant continuants which, when paired, formed units of equal duration. If the temporal relationship is critical, then it would seem necessary to demonstrate how the /mr/ and /nr/ clusters were unique from other experimental pairings. An acoustic explanation for this uniqueness may rest with the assumption that these components are more easily compressed into a type of diphthongized unit. A functional grouping of the nasals with the stops (rather than the continuants) might allow for just such a compression. This regrouping, which is discussed subsequently, would suggest not two continuants, but an instantaneous plosive-like burst of the nasal in conjunction with the vowel-like /r/. Such a two-sound complex, unique in this all-continuant experiment, could be presumably of shorter total duration and consequently more readily-accepted perceptually.

The concept of the nasal grouped with the stops has some support in the literature. Liberman (51) classed the nasals "tentatively" with the stops on the bases that (a) the total duration of the transition is quite short, and (b) there is need of a silent interval between the locus (characteristic frequency position) and the start of the transition. Malécot (61) interpreted his results as support for the grouping of nasals with stops. He said, "Although they (nasals) can be pronounced in isolation without vowels, they generally cannot be separated from them if, perceptually, they are to preserve their phonemic individuality."

Assuming that nasals do behave as stops, what acoustic evidence is there to support the almost instantaneous perception of the nasal/r complex? Results of Liberman et al (53) showed that stop plosives when paired with a vowel result in very rapid perception. The relative time of the first and second formant onsets required to give 75% correct judgments with the voiced and unvoiced stops (/do/ versus /to/) was found to be slightly less than 12 msec. This figure compares very favorably with the results of the m/r and n/r conditions, suggesting, perhaps, an /mr/ and /nr/ perceptual unit. Malécot (61) came to a similar conclusion when he said, "At least where nasal consonants are concerned, the smallest identifiable utterance appears to be the syllable."

Consideration is now given to the perceptual processes functioning to accept the /mr/ or /nr/ cluster as a single unit. A number of investigations have been conducted to study the functions of learning and memory together with the consequential role of the perceptually-familiar unit. Miller, Bruner, and Postman (65) investigated the general issue of contextual constraint in visual areas, noting that in vision as well as in audition one constellation of phonemes may be perceived correctly where the same sounds in a different, less familiar pattern are not. The writers suggested that such results imply that a familiar unit can be reconstructed from a bare minimum of perceptual clues. Malécot (62) showed that a 100-cps steady-state pure tone could be successfully substituted in a perception experiment for a nasal syllabic, (nip + pure tone = "nip 'em"). Brown and Hildum (12), also demonstrated the importance of functional or contextual attributes in speech perception. Their investigation led to the conclusion that when listeners hear speech that is expected to be in the native language, their perceptual identifications are directed by their knowledge of sequential probabilities in the language as well as by the acoustic stimulus.

Now the question arises, does nasal/r (*ner*, *mer*) form a highly probable sequential stimulus? The assumption is made that it does. Unfortunately, proof of such a premise is not available. Fischer-Jørgensen (25) has written that in some languages of the Indo-European type, "the nasals together with /l/ and /r/ may form a functional class of consonants occurring close to the vowel in clusters". Statistical studies by Saporta (79) and Carroll (14) have undertaken frequency analyses of phoneme clusters, but have not tested the nasal/r clusters. Perhaps the greatest statistical support for a "probable cluster" thesis may be found in the work of Harwood and Wright (36), a quantitative study of English word formation based on the Thorndike-Lorge frequency lists. One of their results was that almost one-fourth of the 30,000 words are formed with suffixes added to free underlying words. In the suffix list of 9757 words, the suffix *er* was found to occur 1031 times, second only to *ion* with a count of 1138. The authors listed 36 representative types of *er* suffix words which this investigator analyzed for the preceding consonant. Ten of the 36 examples were *ner* words, the next highest count being five *der*'s. Only one *mer* was tallied. This finding is certainly not offered as proof of high probability of nasal/r clusters, but would lend support to such a premise at least insofar as suffixes are concerned.

These, then, are some possible explanations for the nasal/r results. The acoustic and sequential probability data offer some support for the perceptual unit premise. As for the role of perceptual time, Broadbent (10, p. 271) theorizes, "The central organizing time for highly probable stimuli is supposed shorter than that for improbable ones, and thus the limitation is essentially one of rate of handling information."

THE OTHER PHONEME PAIRS

An attempt has been made to explain the high and low extremes of temporal resolution as reflected in the results of the nasal/r and m/n pairs. Somewhere in between these extremes one might determine a normal resolving time for phonemes of the type herein employed. This investigator believes, however, that such a finding would be simply an arithmetic average which would be a false representation of phoneme resolution in time. The highly complex acoustic nature of the sounds, together with the linguistic sequential probabilities and the influential factors of language learning and retention all militate against a generalization of a common phonemic resolving time.

The remaining phoneme pairs have been grouped into this discussion section for purposes of facilitated handling. No attempt is made to analyze independently, in the manner of previous discussions, the possible reasons for the temporal resolution results. Certainly, previously mentioned factors of masking, acoustic relationships, and linguistic probabilities may be functioning in these stimulus pairs, also. Such explanations cannot be generalized, however, over the entire group of phonemes represented in this section.

Observation of Figures 11 through 16 suggests strongly the role played by inter-onset delay per se. (The partial reversal of performance at 30–50 msec for the r/ð sequence in Figure 12 is an exception to the overall trend.) In conjunction with this aspect, the potential for backward masking should not be neglected. Results of the Elliott (21) study on backward masking indicated rapid changes in the masked threshold when the masking interval, separation between the probe signal and the masking signal, was reduced to the 15- to 0-msec range. Such effects could virtually eliminate the perception of certain stimuli separated from other stimuli by a mere 15 msec, or less.

There is more evidence of the frequency masking described in connection with the r/nasal pairs in the n/ð results, Figure 13. Though the n/ð sequence showed the effects of steadily increasing onset intervals, the ð/n sequence was at or above criterion level for all conditions. The 10-msec ð/n results were actually superior to the 70-msec performance. If one considers /ð/ as the dominant phoneme, its relative dominance, theoretically, becomes less complete as the temporal separation between onsets becomes greater.

In effect, the "oneness" of the /ð/, in this supposition, is probably not questioned at short (10-msec) separations, but gradually as more time intervenes between onsets the possibility of "twoness" becomes more probable. This direction of change is more vividly seen in the overall totals for the n/ð pair, Table II. The same effect may also be observed, but to a lesser extent, in Figure 17, l/r. In this result, the /r/ appears to be the masking phoneme. Once again, reference to the 0-msec boxes of the appropriate phoneme pairs offers clues to the dominant phoneme because of the disproportionate number of selections. In addition, the relation of the resonance frequencies of /r/ overlapping and dominating /l/ may be envisioned through observation of Figure 20. Support for a dominant /r/ in this relationship is suggested through the location of the three energy concentrations found in the so-called speech-hearing frequencies of 500 to 2000 cps. The /l/, on the other hand, is seen to have but two formants, one in the area of 300 cps and the second partially paralleling the third formant of /r/.

Particularly interesting results, perhaps experimental artifacts, are apparent in the v/r and v/l performances, Figures 9 and 10. The combination of high resolution at short onset disparities (10 msec) and the almost identical overall results for both pairs are phenomena which, on the surface, at least, cannot be accounted for as were the circumstances surrounding the excellent performance on the nasal/r pairs. It is obvious that /v/ cannot be labeled and treated as a stop consonant as were the nasals. There is, nevertheless, an observation about these phonemes which might offer some explanation. In words involving either the /v/ and /l/ or /v/ and /r/ the two consonants seem to serve equally well as the initial or final phoneme of a CVC syllable containing any one of a number of intermediate vowels. For example:

valentine	lavender	review	verb
veal	leave	river	varsity
volley	live	rove	very
cavalry	love	rave	virus
vault	elevate	rival	veer

Such a consonant juxtaposition certainly is not unique in English, but may be so insofar as the phonemes of this experiment are concerned. This phoneme relationship could explain the good resolution scores obtained for both sequences of both pairs at short onset differences. It might also be speculated that the susceptibility of such words as *elevate*, *cavalry*, and *relevant* to metathesis (*evelate*, *calvary*, and *revelant*) could be the result of the probability of /v/ and /l/ (or /v/ and /r/) assuming either position in relation to the vowel. Perhaps the very fact that these phonemes may be perceived and ordered when close together in time increases the likelihood of transposition. Support for such assumptions would appear to be dependent upon the validity of the premise that the two members of the two phoneme pairs do assume such positions in relation to one another in significant numbers of words.

The term, "perceptual distinctiveness", describes perhaps better than any other expression the complexities apparent in discriminating and ordering phonemes. Fischer-Jørgensen (25) considered this aspect as it relates to the time continuum in his paper delivered at the 1952 Conference on Speech Analysis. He said:

> The general character of phonemic systems, e.g., the tendency to have relatively great phonetic distances between phonemes and to have a certain symmetry in the pattern (particularly obvious in stops and vowels), have been explained by Roman Jacobson from the point of view of perception: it is easier to keep phonemes apart when they are perceptually very different, and easier to recognize a limited number of features. It might perhaps equally well be maintained that it is easier to keep different articulations apart and to use the same distinctive features in different pairs.

Carroll's (14) conclusion from his statistical study of words was that there are significant trends in favor of the hypothesis that clustering is somehow dependent upon the distinctive features of phonemes.

To illustrate this perceptual distinctiveness, an example may be drawn from the present investigation. Since results for the v/r were almost identical to those of the v/l, it is logical to expect a close similarity in the performances of m/r and m/l. Such was not the case, however. The m/r pair was temporally resolved more accurately than any other stimulus pair in the experiment, whereas the m/l results of the second pilot investigation were not sufficiently good to qualify the stimulus pair for inclusion in the major study. In their speech synthesizing studies, Cooper et al. (18) found that a transition from higher to lower frequency which is followed by a steady-state resonant sound is often heard as /m/ but may at times sound like /l/ instead. The point to be made here is that perceptual distinctiveness is highly influenced by the phoneme environment, not simply the characteristics of the individual phoneme. By creating a time-controlled environment in this study, the importance of this relationship is made apparent.

THE PURE-TONE PAIR

The point was made in Part I that Hirsh (41), using a variety of tones, noises, and clicks, and Broadbent and Ladefoged (11), employing combinations of non-verbal *pip*, *hiss*, and *buzz*, generalized their temporal findings to speech. Hirsh, who used trained subjects, found that the time intervening between two sounds that was required for 75% correct judgments was of the order of 20 msec and was independent of the kinds of sounds used. For pure tones Hirsh obtained a 17-msec resolution. Broadbent and Ladefoged concluded that originally large (150-msec) resolution times by naive listeners could be reduced to the magnitude found by Hirsh after repeated exposure. Kinney (48) found that pairs of non-pure, pattern playback tone spectrograms could be resolved 75% correctly after 30-40 msec. Results of the present investigation indicate that untrained subjects could resolve pure tones of 1200 and 250 cps with 75% accuracy in the 30- to 50-msec range of onset separation. Presumably, subject training could have resulted in approximately 20-msec discriminations for both the Kinney study and the present investigation.

The important point to be noted, however, is not the comparison of non-verbal results, but the comparisons of non-verbal results with those obtained from the ordering of speech samples. Ordering (for example, m/r) was above 75% accurate for untrained subjects at onset disparity of but 10 msec. The conclusion may be drawn that normal adults are not "untrained" when their native language is the stimulus.

The results of Liberman et al (53) in their study resolving differences between /do/ and /to/ by successively cutting back the first formant in relation to the second formant of their painted spectrogram show that 75% correct labeling judgments were obtained at "slightly less than 12 msec". They concluded that the discrimination peak at the phoneme boundary was a result of learning. Their results with stop consonants compare favorably with the best results obtained in the present investigation. It seems that learning and memory are indeed operating in speech perception investigations.

One final observation must be made. In discussing the results of Hirsh (41), Broadbent and Ladefoged (11) concluded that the Hirsh findings were dependent in part upon the quality of the two-sound complex rather than upon differences of perceived order. At least insofar as understanding speech perception is concerned, this multi-phoneme quality seems of vital importance. The concepts of environmental phoneme relations, overlap, sequential dependency, and perceptual segmentation appear to rest upon this very fact. Silversten (82) has described in the context of synthetic speech the value of thinking and synthesizing in terms of time segments of varying size and type: phonemes, phoneme dyads, syllable nuclei and margins, half-syllables, syllables, syllable dyads, and words. The instrumental technique employed in the present investigation may be of value for analyzing time segments larger than the single phoneme.

SUMMARY

Results of this investigation should be considered with the following experimentally imposed limitations in mind: (a) Phoneme samples were of one speaker and were relatively invariant in nature, lacking both normal rise- and decay-times; (b) all sound stimuli were recorded at equal record-level as determined by VU meter monitoring; (c) listening was exclusively monaural; and (d) choices were forced from two provided options.

The general findings may be summarized as follows:

a. Temporal resolution (as reflected by criterion median scores of 75% or better) varied according to the phonemes involved. Performance ranged from high accuracy at short (10 msec) onset disparities for m/r, n/r, v/l, and v/r, to failure to meet criterion level even at 70-msec inter-onset disparity for m/n.

b. Temporal resolution often varied according to the particular phoneme of a stimulus pair that was presented initially. Performances on sequences z/r, ð/r, n/z, l/z, l/ð, for example, were better than for the reverse onset sequences r/z, r/ð, z/n, z/l, and ð/l.

c. When viewed from the phoneme classification frame of reference, the results revealed both inter-classification similarities and differences in temporal resolution. The two nasals, /m/ and /n/, paired with semi-vowel /r/ gave similar results as did the two semi-vowels, /r/ and /l/, when paired with fricative /v/. On the other hand, fricatives /z/ and /ð/ paired with semi-vowel /l/, though showing similar results to one another, were quite different from the /v/ plus semi-vowel results. Pairings from within phoneme classifications, r/l and m/n, were poorly resolved even under conditions of relatively long onset intervals.

d. Temporal resolution of a pure-tone pair, 1200/250 cps, was inferior to that found for the m/r, n/r, v/l, v/r, z/r, and ð/r phoneme pairs.

e. Temporal resolution of two members of a sound pair beginning simultaneously varied with the sounds involved. Selections for one sound over another ranged from proportions of 13%-87% to 48%-52%.

PART III

THEORETICAL IMPLICATIONS

"... The acoustic facts do not help if they are not seen either as a result of articulation or as a cause of perception", wrote Fischer-Jørgensen (25), "for speech is human activity, and explanations must be found in human physiology and psychology." In these concluding pages an attempt is made to view the experimental findings of Part II together with the pertinent information of Part I in the light of human physiology and psychology.

What appears to be the critical factor in any speech perception experiment is the listener's reference patterns, or the so-called "comparator" of servo models. Man's comparison patterns for speech are language patterns and, regardless of whether storage is predominantly auditory or motor in nature, the essential dynamics of language itself cannot be overlooked. Our linguistic code has evolved from what Zipf (99, p. 95) terms the "accidents of arrangement of vocal apparatus". Irrespective of semantic value, one may presume that most words which in their linguistic inception and diachronic development have not molded in conformance to the imposed limitations of the speaking apparatus have failed to survive. One of the restrictions forced upon man's linguistic utterances is the relative ease or economy of phonatory and oral articulatory patterns. Such criteria are factors, no doubt, in much linguistic change and could suggest why metatheses of pure motor character can become permanent linguistic alterations. The word (or its meaningless components) is, therefore, the common denominator of comparison patterns. Contrary to inferred non-verbal perceptual processes, the phonemic pattern is the essence of comparison and the ensuing processes of speech perception and production. This seemingly understated premise all too often is largely, if not totally, ignored in the predominantly acoustical analyses of auditory perception. Most of the previously cited temporal sequence investigators have generalized or inferred speech perceptual behavior from acoustic stimuli which, despite many such qualities as finite controllability, remained non-verbal.

In an experiment of the type reported in Part II of this book, the adult listeners had at least fifteen years of speaking and listening experience with its consequential build-up of linguistic storage of comparison patterns. They were, in a very literal sense, extremely well-practiced observers in their assigned listening experience – but only for the phonemic comparisons, not for the pure tones. Although temporal ordering of acoustic signals is essential to some specialized skills such as telegraphy and sonar, considerable specialized practice is a necessary prerequisite to adequate performance. Suffice to say, people are seldom required to impose an order upon noises or tones, the predominant temporal characteristics of which might better be described in terms of duration, rhythm, or succession. Therefore, the subjects' performance on the pure-tone stimuli, despite the acoustic simplicity thereof, is explainable. That one half of the phoneme pairs were better ordered temporally suggests strongly the advantage afforded by fifteen or more years of linguistic comparator experience. It will be recalled that Hirsh's (41) highly practiced subjects resolved the same similarly presented pure tones with 75% accuracy at 17 msec separation, which was even then at least 7 msec greater in ordering time than the best of the phonemic pairs. It may also be remembered that in a replication study, Broadbent and Ladefoged (11) found that their initially long thresholds of resolution were reduced to the magnitude found by Hirsh after repeated exposure. They concluded that "listeners have to be trained to interpret as order those cues which the ear transmits about the relative arrival of the stimuli".

Now it might be argued that the experimental phonemic stimuli, because of (a) their extraction from prolonged utterances, (b) their uniform rise and abrupt decay times, (c) their standard VU-level recording, and (d) their monaural presentation, were not, in fact, truly representative of the acoustic character of contextual verbalization of voiced consonant continuants. While this observation may or may not be valid, the essential fact remains that each listener had both a prescribed indication of the component phonemes of each stimulus pair and a linguistic comparison potential established through years of language usage. Therefore, the preparatory set with which each listener would approach his task would seem to be quite different from that anticipatory to a high/low pure-tone discrimination. It may be recalled that Brown and Hildum (12) stressed the importance of expectation of one's native language upon perceptual results. The experimental data suggest strongly that to the extent the phonemic elements either shared a perceptual distinctiveness or comprised a familiar linguistic cluster they could be temporally ordered with shorter onset differentials than required for similar non-verbal stimuli. In brief, speech is uniquely man's, and perceiving speech is unique in man's listening experiences. This difference was so vividly expressed by Cherry (16, p. 293) in these words, "A child learns to imitate the speech sounds of its mother; it does not learn to make sounds like bells or frying bacon."

Evidence of this study is seen to support a general assumption that the perceptual ease with which articulations of the sound wave can be handled in time is contingent upon the interdependency of speech perceptual and articulatory processes and their consequential by-product, language configurations. Sequential dependencies of the resultant language, in turn, contribute to expeditious perception. The limitations imposed upon the word in its diachronic development may be very similar to the motoperceptual limitations imposed upon the young child's word in his linguistic development until such time as the word(s) (or components thereof) becomes a dynamic force in perception.

What might be termed a "theoretical footnote" to the contents of this volume will conclude it. Support for the perceptual inferences to be made is not forthcoming from the experimental investigation although the study of both Parts I and II suggested strongly the need for reconsideration of conventional speech perception theories in light of a linguistic comparator system handling sequential speech.

Speech is motor behavior with an acoustical product. That both motor and sensory elements share in the perceptual process has been theorized by Licklider (55) and Fischer-Jørgensen (26) among others. Other authorities tend to endorse strongly the role of one perceptual avenue or the other. For example, Peterson (73) maintains that learning to speak requires comparison of remembered signals with those of others. This memory, he believes, cannot be within the motor controlling portions of the brain because acquisition of sensory patterns must precede the acquisition of corresponding motor productions. On the other extreme, Liberman (51) maintains that "where articulation and sound wave go separate ways ... the perception always goes with articulation". Twaddell (90) suggests that a hearer matches the acoustic stimuli he receives against his own habits of muscular speech action, and identifies the incoming sound as corresponding to his own speech articulations. "At both ends of a speech transmission", Twaddell states, "it is muscular activity, not the acoustic character, which dominates the identification."

Could it be this muscular activity or its mental representation which provides the uniqueness of the speech listening process from that of other listening? Does the preparation or "mental set" preliminary to receipt of speech have a muscular character? Could such a preparatory set be reflected in the *forward dependency* factor eluded to previously? Certainly, the concept of motor readiness in human perception has been extensively theorized and investigated. (For an account and critique of motor set in perception, see Allport [1, Chapter 8].) And the attractiveness of theories of perception such as offered by Sperry (84) certainly would not exclude the speech act. His theory, based upon phylogenetic and neurologic considerations, maintains that motor adjustment, rather than stimulus patterns or the contents of subjective experience, figures predominantly as a proper frame of reference for understanding the organization, meaning, and significance of brain excitation. "Temporally as well as spatially, the mental and motor patterns must integrate, mesh, and interlock." Perhaps this describes the theoretical process whereby speech flow is segmented into units.

Now there is little question that perception of speech relies upon some amount of auditory processing of the acoustical signal. Nevertheless, it is suggested that a major, if not predominant, analyzer is motor adjustment, at least in regard to consonant perception. That the two processes operate as a binary system to "integrate, mesh, and interlock" the elements of speech is strongly suggested in the metathesis phenomenon as well as in numerous clinical observations of verbal output when one of the processes has become faulty. Might not the anticipatory and non-contiguous character of clinical metathesis reflect the integrative failure of the motor system in the juxtapositioning of the consonants about a constant sensori-associative vowel system? Failure of auditory memory, if such a system exists, fails to account for *retention* of the involved consonants and does not readily explain an integrative attribute of short-term storage. Temporal sequence confusion of the motorium would possibly explain the character of metathesis in both its predominantly motor (linguistic metathesis) and its motor and/or perceptual-associative (clinical metathesis) forms. Motor-based temporal confusion could be imposed by the word configuration, itself, or by neuromuscular limitations as reflected in perception and/or oral production. Such limitations in the verbal-motor operation of an associated skill may be visualized in the performance of novice typists whose extensive metathesized output is anticipatory and very often reflected as juxtapositioned consonants about a typed vowel.

The potentially greater influence of the motor contributions to the perceptual act has been largely ignored in speech studies by what Sperry terms "our present one-sided preoccupation with the sensory avenues to the study of mental processes". Deaf education, for example, might assume a quite different perspective if supportive research indicated the involvement of articulatory kinesthesia in the verbal perception process. Speech reading might then include an active serial motor component. Similarly, verbal auditory memory might be more successfully investigated as audio-articulatory memory.

To summarize, the characteristics of language per se, together with the motor components embodied in the language unit and in its oral output, suggest the extensive role which motor systems of the brain might also play in extra-auditory perception and learning of oral language. Time assumes a vital psychophysical role in processing the sequential character of language receipt. When temporal ordering performance of language exceeds in many instances similar performance with non-verbal stimuli, then the potential for superiority would seem to rest in the verbal learning and verbal comparator operations. There is some evidence to suggest that these operations, in turn, are facilitated through the probability of familiar phonemic clusters.

APPENDICES

APPENDIX A

THE PILOT INVESTIGATIONS

The instrumental methods and rationale for the procedural techniques used in the two pilot studies are described in detail in the body of this book and are therefore not repeated in the following brief accounts of the preliminary investigations.

Pilot Study I

The purposes of the initial pilot study were to investigate the feasibility of the instrumental technique, to determine the general experimental procedure, and to establish the type and range of data obtained from the listening judgments.

With the experimenter serving as speaker, tape recorded samples were made of sustained utterances of /v/, /l/, /ð/, /r/, /z/, and /m/. The basis for selection of the phonemes was that they be consonants of the continuant type and therefore suitable for sustained prolongation. This was necessary for the subsequent tape-splicing procedure used to equate rise-times. In addition, to evaluate the procedural method with that employed by Hirsh (41) in one of his experiments, recordings were made of two pure tones, 1200 and 250 cps. Tape speed was 7½ inches per second (ips), and all stimuli were monitored at a constant VU level. The stimuli were subsequently paired and mixed in such a way that the onsets of the members could be precisely timed in relation to one another. The result was a master magnetic tape with the following stimulus pairs represented: v/l, ð/r, z/m, 1200/250. Onset-disparities were –60, –40, –20, 0, 20, 40, and 60 msec. Ten samples of each onset-disparity were recorded on the master tape for each of the four stimulus pairs. For example, the first ten stimuli were the v/l pair separated in onset time by 60 msec, the second ten stimuli were v/l separated by 40 msec, and so on through all temporal conditions. (It should be noted that this condition sequence was abandoned for the major investigation. A randomization procedure was substituted in order to break up the continuity here described.) The component sounds were mixed following the onset of the second, and the resultant complex was as long as 500 msec duration, the length of the longest member. A period of 4.5 seconds separated each of the 280 stimulus samples of the master tape.

Subjects were 12 Purdue University undergraduate females ranging in age from 19 to 21 years and drawn from an introductory methods course in speech correction. Each girl was without history of hearing pathology and passed an audiometric screening test of the right ear. The subjects, seen individually, were seated in a sound-treated booth and fitted with a headset, the left phone of which was inactive. Instructions were to listen to each presentation and then to mark on a sheet provided which of the two sounds indicated (for example, "V" or "L") was heard first. Responses were to be made to each presentation, guessing if necessary. The presentation order of the sound pairs was randomized for each subject, but the onset-disparities were in constant sequence for all subjects. The master tapes were delivered to the right headphone at a level of 65 db SPL (re .0002 microbar) from the tape reproducer located outside the subject's booth. (A different listening environment for the major investigation necessitated a change from the level herein employed.) Rest periods were provided after each of the four sound-pair conditions.

The results of this investigation are graphically presented in Figure 21 of this appendix. The percentage of correct responses of each subject for each experimental condition was averaged over all subjects. The 75% level, shown on the figure, was used as the criterion score to indicate correct performance. It will be noted that there was considerable variability both among the phoneme pairs, and as a result of reversed initial onset. It was decided from these results that the temporal displacement range was not sufficiently great (either toward or away from zero) to fully evaluate the effects of temporal onset manipulation.

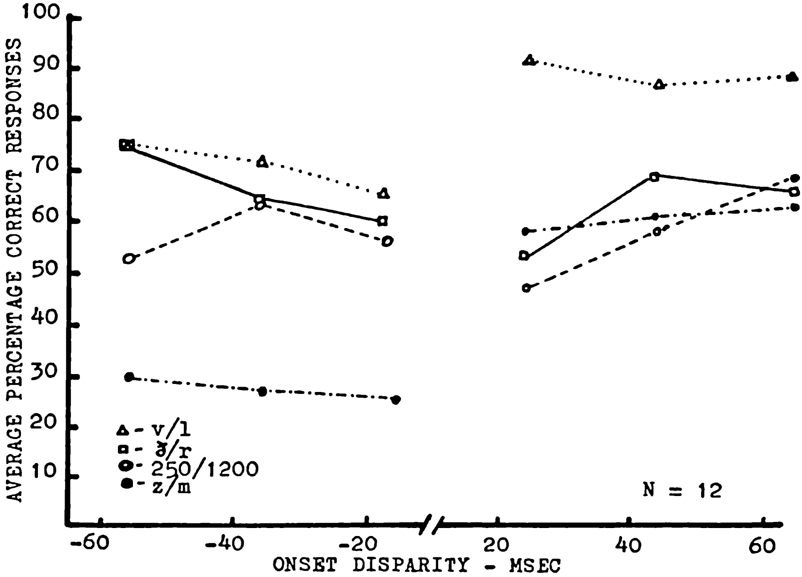


Figure 21. Graphical representation of the results of Pilot Study I. Average percentage of correct identification of first element in a stimulus pair.

The pure-tone results agreed in part with Hirsh's (41) findings. The performance curve for this condition in Hirsh's study was, however, somewhat steeper, indicating more rapid criterion (75%) achievement.

Among the data analyzed were the average correct responses over all subjects for the *first* presentation of each new listening condition. Figure 22 compares these correct first responses with the average correct responses for each ten trials by stimulus pair conditions. It may be noted that the 10-item average results in about a 5% better score than the first-item results for all conditions excepting ð/r in which the first-item mean was superior.

Pilot Study II

The purpose of the second pilot study was to collect data which would aid in the selection of the phoneme pairs to be used. The

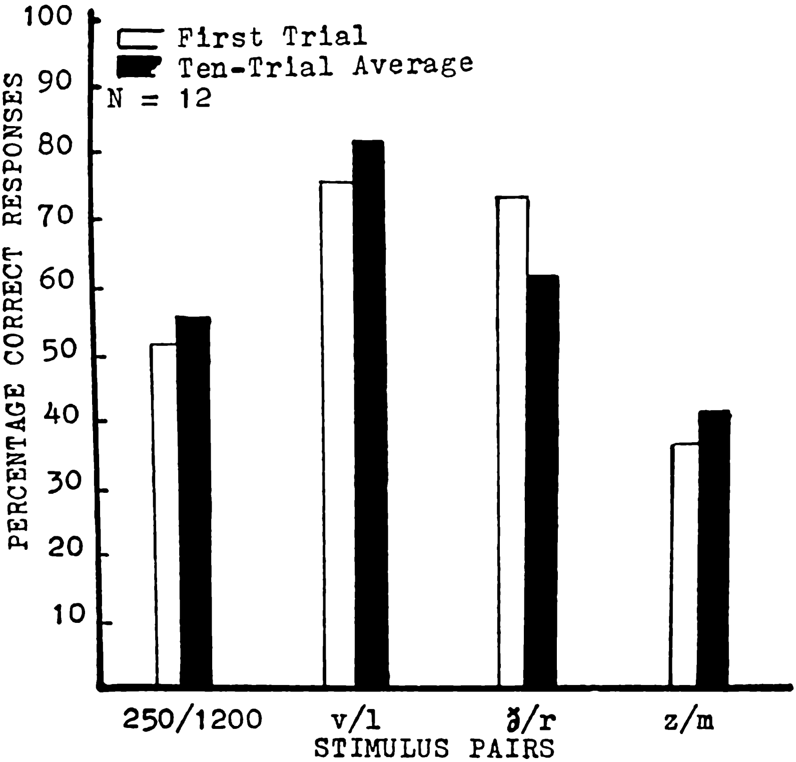


Figure 22. Comparison of correct responses of Pilot Study I, first presentation and ten-trial average.

selection was limited to phonemes of the voiced consonant continuant classification for reasons previously stated. With the exception of the four sound pairs used in the first pilot study, all possible combinations of the following phonemes were recorded and subsequently paired: /v/, /l/, /z/, /m/, /n/, /r/, and /ð/. (The Pilot Study I pairs were excluded because performance on these pairs was determined.) The recording and timing procedures were essentially the same as those used in the earlier and later studies except that (a) only two onset-disparities were employed for each stimulus pair, –40 and 40 msec, and (b) only four samples of each temporal condition were recorded. The order of appearance of the 18 pairs at two onset times (36 conditions) was randomized.

Six subjects, five females and one male, were seen individually as in the first pilot investigation. All subjects were Purdue speech pathology majors ranging in age from 20 to 25. These listeners represented what was believed to be a slightly higher level of sophistication than those of Pilot Study I insofar as listening discrimination is concerned. The subjects were, however, naive to the experiment, and their potential for making more accurate discriminations was considered advantageous in the light of the needs of this particular study. The room and apparatus set-up of the first pilot study were again employed. After audiometric screening the subject was given a pre-marked sheet on which a sound pair was indicated for every four item blanks on the paper. (The sequence of the sound pairs involved was determined by the randomizing process.) Instructions again included a request to indicate the first-heard sound after presentation of each pair. For this study, however, the subject was instructed to leave an item blank if he found the decision impossible to make. (This is a departure from the forced-choice methods used in both the first pilot study and the major investigation. It was used here because of the purpose of this study, not to force temporal resolution, but rather to ascertain on which pairs temporal resolution was obtainable. Theoretically, a forced-choice procedure could have resulted in misleading data due to chance performance when but four responses were requested and but two onset-disparities were used.)

The results of Pilot Study II are shown in Figure 23 of this appendix. The total correct of 48 responses was averaged over all subjects. The 75% level is indicated, as are the converted $\pm$ 40-msec scores of the first pilot study.

Among the decisions resulting from this study are the following:

a. All fricative/semi-vowel pairs were to be included in the major investigation because of the general superior performance on these phonemic classifications.

b. The pure-tone pair was to be included for comparison potential with the phonemes and for reasons of the design re-

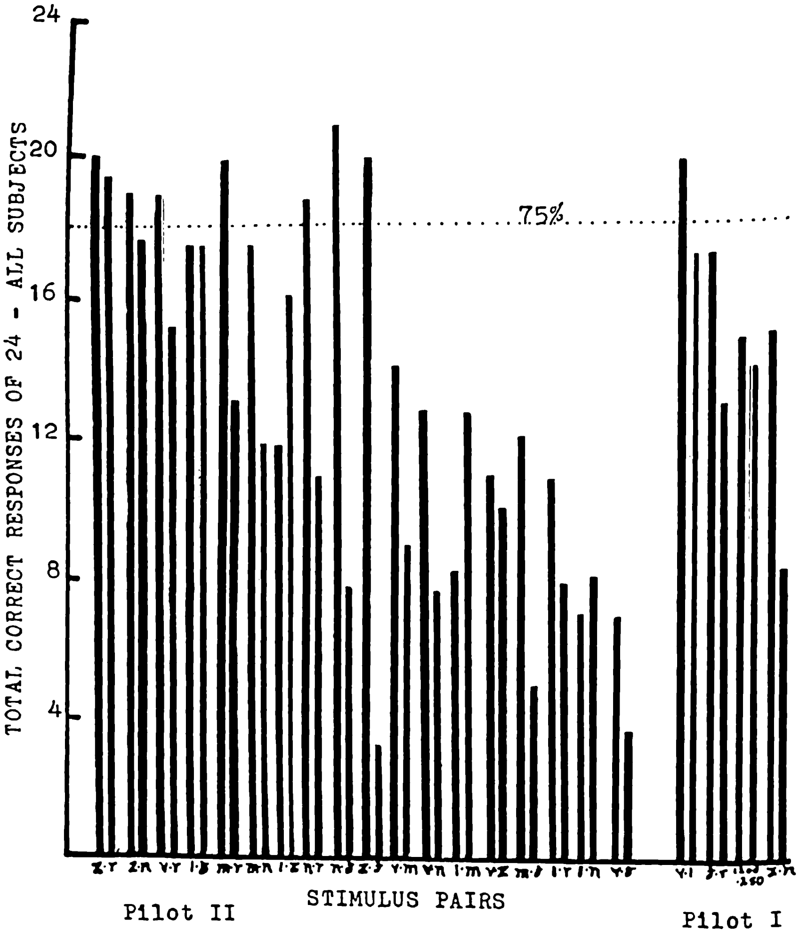


Figure 23. Results of the 40-msec onset separations of Pilot Study II together with the 40-msec results for the pairs of Pilot Study I. The 75%-level of performance is indicated. Each bar represents performance when the phoneme indicated beneath preceded in onset the associated member of the pair.

quirement of a constant performance criterion over all subject groups.

c. The consistently inferior performance on the paired fricatives led to the conclusion that these pairings were unsatisfactory for an experiment involving the relatively narrow temporal range to be employed. They were therefore excluded.

d. Although performance on the m/n and r/l pairs was not high, these pairs were endorsed for inclusion because they represented two unique combinations of pairings within the same phonemic classifications. The m/n pair would have been excluded because of the potential for discrimination difficulty (as was true with the paired fricatives) were it not for the fact that this pair yielded the best results of any within-phoneme classification pairings.

e. The remaining phoneme pairs were included if combined 40 and –40 msec scores were at or above an arbitrarily-assigned criterion of 28.

APPENDIX B

DESCRIPTION OF MAJOR INSTRUMENTATION

This section describes more fully the major instrumentation employed. Descriptions appearing in the main body of the book are excluded.

Ampex Tape Recorder, Model 300-2C

This two-channel recorder allows high fidelity operation at two speeds. Signal strength may be monitored by independently-mounted VU meters. For this experiment a tape speed of 7½ ips was used exclusively, and all two-channel record and playback procedures were monitored and controlled for equal and constant VU levels.

Ampex Tape Recorder, Model 601

This instrument is a single-channel high fidelity system. It is portable and operates at a pre-set single speed of 7½ ips. The recorder is equipped with a VU meter.

Tektronix Type 502 Dual-Beam Oscilloscope

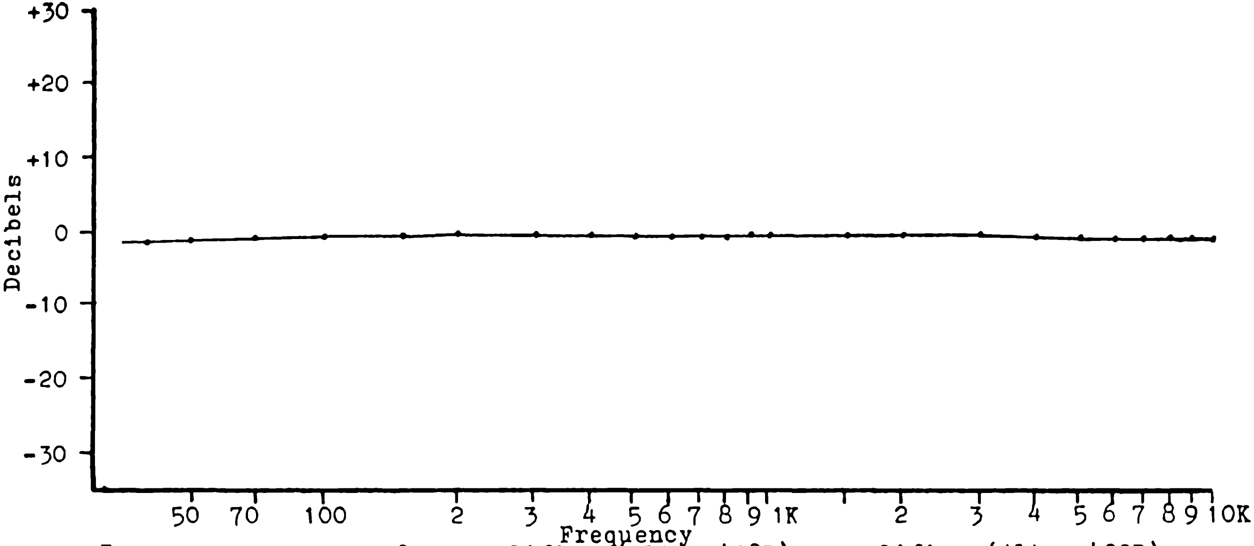
This instrument may be used to compare and measure the outputs of two transducers on the same time base. Sensitivities at all positions are within 3% of the panel readings. For this experiment a calibrated sweep rate of 2.0 msec/cm was used, and triggering was external.

Apparatus: Measured Specifications

	S/N Ratio	*Wow*	*Flutter*	*Harmonic Distortion*
Preamp -Altec 428B plus Amp Altec 429B	70 db	—	—	0.139%
Recorder Ampex 300 Channel 1*	42.5 db	0.03%	0.06%	0.5%
Recorder Ampex 300 Channel 2*	42.0 db	0.05%	0.07%	0.8%
Recorder Ampex 601	35.0 db	0.06%	0.2%	0.9%
Echo-Vox Channel 1 20-540 msec	35.7 db	0.12%	0.05%	4.72%

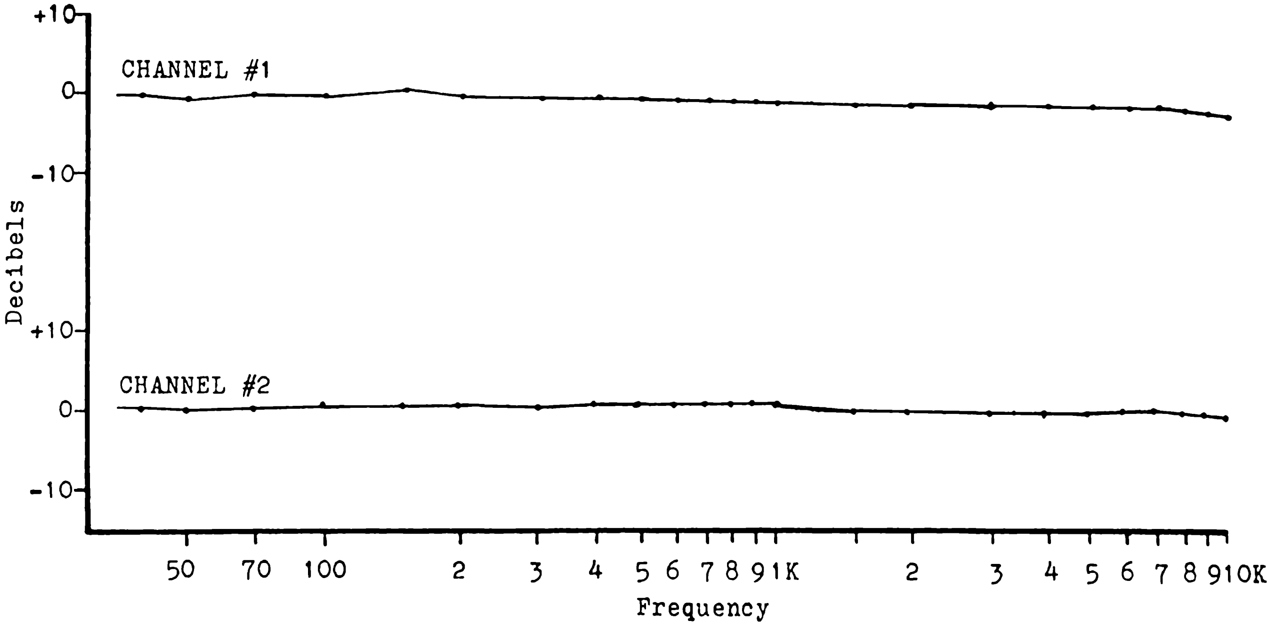
* Ampex 300 Channel Separation (Ch. 1 to Ch. 2): -40 db.

APPENDIX B



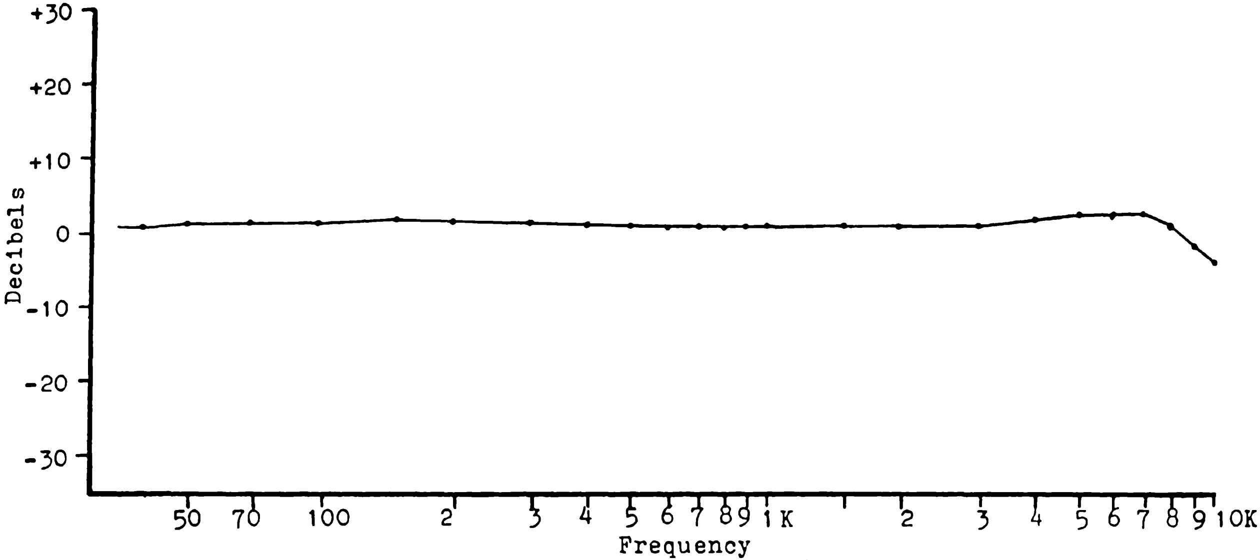
Frequency response of preamplifier (Altec 428B) + amplifier (Altec 429B)

APPENDIX B

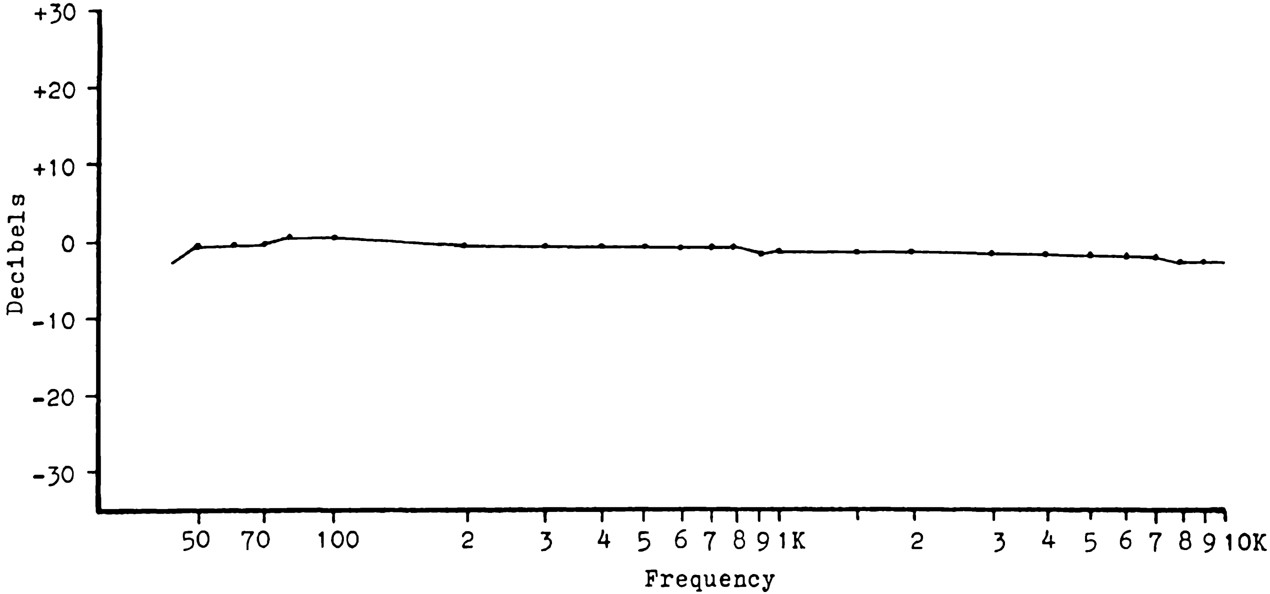


Frequency responses of the two channels of the tape recorder, Ampex 300.

APPENDIX B

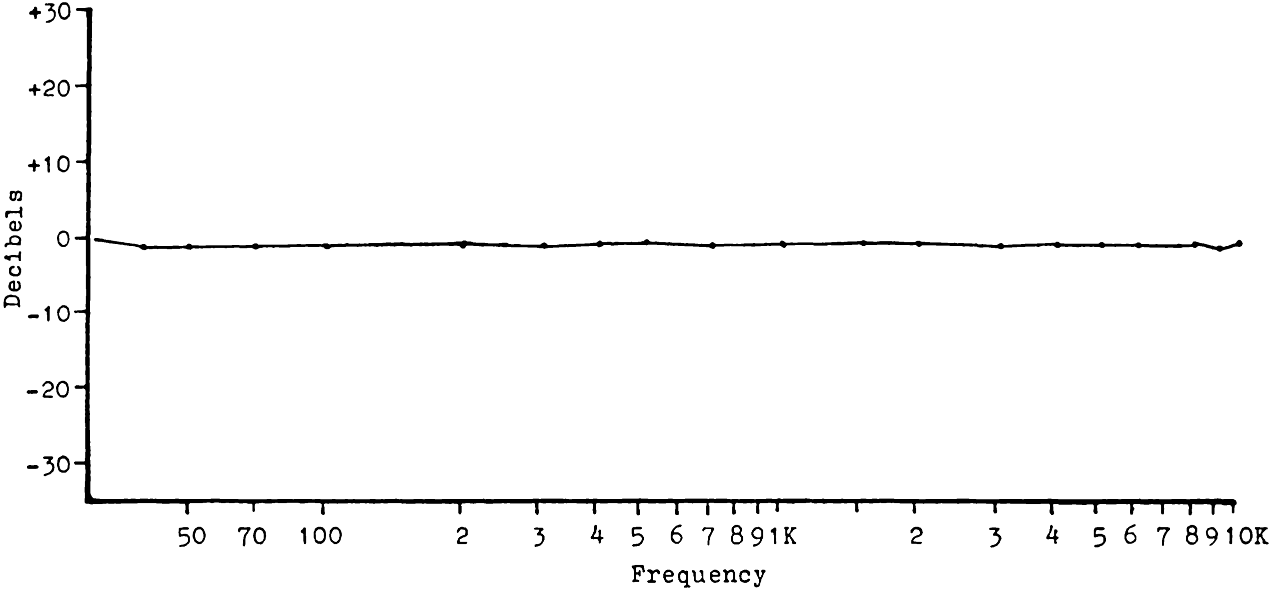


Frequency response of the tape recorder, Ampex 601.



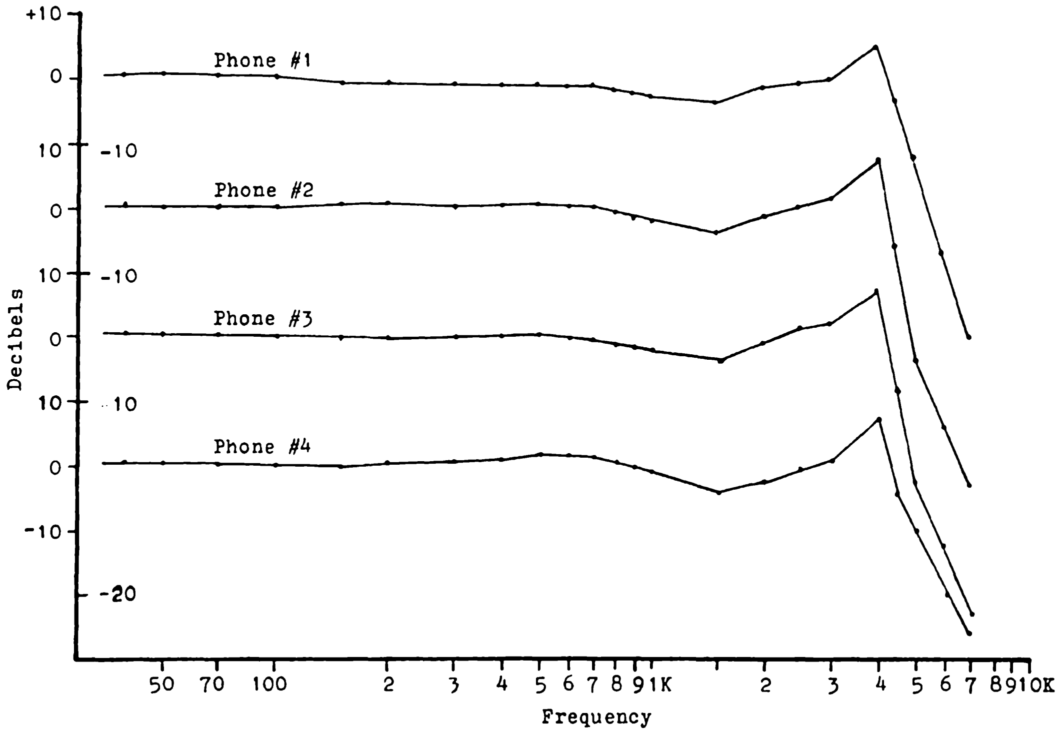
Frequency response of Channel #1, Echo-Vox Sr., multiple time delay unit.

APPENDIX B



Frequency response characteristics of the balancing transformer, UTC LS-33.

APPENDIX B



Frequency response charactersitics of the four earphones employed, TDH-39.

BIBLIOGRAPHY

1. Allport, F. H., *Theories of Perception and the Concept of Structure* (New York, John Wiley and Sons, 1955).
2. Ammassian, V. and Waller, H. J., "Spatiotemporal patterns of activity in individual reticular neurons", in *Reticular Formation of the Brain* (Boston, Little, Brown, 1958).
3. Arnold, G. E., "Studies in tachyphemia: I. Present concepts of etiologic factors", *Logos*, 3 (1960), pp. 25-45.
4. Attneave, F., *Applications of Information Theory to Psychology* (New York, Henry Holt, 1959).
5. Ax, A. F. and Colley, W. H., "Temporal acuity of vision, audition, and touch in psychogenic and neurogenic pathology", *J. consult. Psychol.*, 19 (1955), pp. 455-461.
6. Babkoff, H. and Sutton, S., "Perception of temporal order and loudness judgments for dichotic clicks", *J. acoust. Soc. Amer.*, 35 (1963), pp. 574-577.
7. Backus, O. L., *Speech in Education* (New York, Longmans, Green, 1943).
8. Berry, Mildred F. and Eisenson, J., *Speech Disorders* (New York, Apple-ton-Century-Crofts, 1956).
9. Black, J. W., "The reception of repeated and overlapping speech patterns", *J. acoust. Soc. Amer.*, 27 (1955), pp. 494-496.
10. Broadbent, D. E., *Perception and Communication* (New York, Pergamon Press, 1958).
11. ——, and Ladefoged, P., "Auditory perception of temporal order", *J. acoust. Soc. Amer.*, 31 (1959), pp. 1539.
12. Brown, R. W. and Hildum, D. C., "Expectancy and the perception of syllables", *Language*, 32 (1956), pp. 411-419.
13. Callaway, E. III, and Alexander, J. D. Jr., "The temporal coding of sensory data: an investigation of two theories", *J. gen. Psychol.*, 62 (1960), pp. 293-309.
14. Carroll, J. B., "The assessment of phoneme cluster frequencies", *Language*, 34 (1958), pp. 267-278.
15. Cheatham, P. G. and White, C. T., "Temporal numerosity: III. Auditory perception of number", *J. exp. Psychol.*, 47 (1954), pp. 425-432.
16. Cherry, C., *On Human Communication* (New York, John Wiley, 1957).
17. Chistovich, L. A., "The significance of temporal factors in the reception of complex sounds", *Soviet Physics – Acoustics*, 2 (1956), pp. 328-334.
18. Cooper, S., Delattre, P. C., Liberman, A., Borst, J. M., and Gerstman,

L. J., "Some experiments on the perception of synthetic speech sounds", *J. acoust. Soc. Amer.*, 24 (1952), pp. 597-606.

19. Davis, H., "Peripheral coding of auditory information", in W. A. Rosenblith (ed.), *Sensory Communication* (New York, John Wiley, 1961).
20. Efrom, R., "Temporal perception, aphasia and déjà vu", *Brain*, 86 (1963), pp. 403-424.
21. Elliott, Lois L., "Backward masking: monotic and dichotic conditions", *Sch. of Aerospace Medicine Technical Documentary Rep., SAM-TDR-62-76*, June 1962 (School of Aerospace Medicine, Brooks AF Base, Texas).
22. Fairbanks, G., "Systematic research in experimental phonetics: 1. A theory of the speech mechanism as a servosystem", *J. Speech Hearing Dis.*, 19 (1954), pp. 133-139.
23. Fano, R. M., "The information theory point of view in speech communication", *J. acoust. Soc. Amer.*, 22 (1950), pp. 691-697.
24. Fant, C. G., "Descriptive analysis of the acoustic aspects of speech", *Logos*, 5 (1962), pp. 3-17.
25. Fischer-Jørgensen, E., "The phonetic basis for identification of phonemic elements", *J. acoust. Soc. Amer.*, 24 (1952), pp. 611-617.
26. ——, "What can the new techniques of acoustic phonetics contribute to linguistics?", in S. Saporta (ed.), *Psycholinguistics: A Book of Readings* (New York, Holt, Rinehart and Winston, 1961).
27. Fletcher, H., *Speech and Hearing in Communication* (New York, D. Van Nostrand, 1953).
28. Frick, F. C., "Some perceptual problems from the point of view of information theory", in *Current Trends in Information Theory* (Pittsburgh, U. of Pittsburgh Press, 1953).
29. Fry, D. B., "Phonemic substitutions in an aphasic patient", *Lang. Speech*, 2 (1959), pp. 52-61.
30. Furth, H. G., "Sequence learning in aphasic and deaf children", *J. Speech Hearing Dis.*, 29 (1963), pp. 171-177.
31. Garner, W. R., "The accuracy of counting repeated short tones", *J. exp. Psychol.*, 41 (1951), pp. 310-316.
32. Goldberg, I., *The Wonder of Words* (New York, D. Appleton-Century, 1938).
33. Gray, G. W. and Wise, C. M., *The Bases of Speech*. Third Edition (New York, Harper, 1959).
34. Guttman, N., van Bergeijk, W. A., and David, E. E. Jr., "Monaural temporal masking investigated by binaural interaction", *J. acoust. Soc. Amer.*, 32 (1960), pp. 1329-1336.
35. Hardy, W. G., "Problems of audition, perception and understanding", *Volta Rev.*, 58 (1956), pp. 289-300, 309.
36. Harwood, F. W. and Wright, A. M., "Statistical study of English word formation", *Language*, 32 (1956), pp. 260-273.
37. Heise, G. A. and Miller, G. A., "An experimental study of auditory patterns", *Amer. J. Psychol.*, 64 (1951), pp. 68-77.
38. de Hirsch, Katrina, "Prediction of future reading disabilities in children with oral language disorders", *Folia Phoniatr.*, 7 (1955), pp. 235-250.

39. Hirsh, I. J., "Certain temporal factors in audition", *Science*, 116 (1952), pp. 523-524.
40. ——, *The Measurement of Hearing* (New York, McGraw-Hill, 1952).
41. ——, "Auditory perception of temporal order", *J. acoust. Soc. Amer.*, 31 (1959), pp. 759-767.
42. ——, and Sherrick, E. E. Jr., "Perceived order in different sense modalities", *J. exp. Psychol.*, 62 (1961), pp. 423-432.
43. Hockett, C. F., Review of *The mathematical theory of communication* by Claude L. Shannon and Warren Weaver, in *Language*, 29 (1953), pp. 69-93.
44. Huggins, W. H., "System-function analysis of speech sounds", *J. acoust. Soc. Amer.*, 22 (1950), pp. 765-769.
45. ——, "A phase principle for complex-frequency analysis and its implication in auditory theory", *J. acoust. Soc. Amer.*, 24 (1952), pp. 585-589.
46. James, W., *Principles of Psychology* (New York, Holt, 1890).
47. Joos, M., "Description of language design", *J. acoust. Soc. Amer.*, 22 (1950), pp. 701-708.
48. Kinney, Jo Ann S., "Discrimination in auditory and visual patterns", *Amer. J. Psychol.*, 74 (1961), pp. 529-541.
49. Ladefoged, P. and Broadbent, D. E., "Perception of sequences in auditory events", *Quart. J. exp. Psychol.*, 12 (1960), pp. 162-170.
50. Lashley, K. S., "The problem of serial order in behavior", in S. Saporta (ed.), *Psycholinguistics: A Book of Readings* (New York, Holt, Rinehart and Winston, 1961).
51. Liberman, A. M., "Some results of research on speech perception", *J. acoust. Soc. Amer.*, 29 (1957), pp. 117-123.
52. ——, Harris, Katherine S., Eimas, P., Lisker, L., and Bastian, J., "An effect of learning on speech perception: the discrimination of durations of silence with and without phonemic significance", *Lang. Speech*, 4 (1961), pp. 175-195.
53. ——, Harris, Katherine S., Kinney, Jo Ann, and Lane, H., "The discrimination of relative onset-time of the components of certain speech and nonspeech patterns", *J. exp. Psychol.*, 61 (1961), pp. 379-388.
54. Lichtheim, L., "On aphasia", *Brain*, 7 (1885), pp. 433-484.
55. Licklider, J. C. R., "On the process of speech perception", *J. acoust. Soc. Amer.*, 24 (1952), pp. 590-594.
56. ——, "Hearing", *Annu. Rev. Psychol.*, 4 (1953), pp. 106-107.
57. ——, and Miller, G. A., "The perception of speech", in S. S. Stevens (ed.), *Handbook of Experimental Psychology* (New York, John Wiley, 1951).
58. Lindsley, D. B., "Psychological phenomena and the electroencephalogram", *EEG Clin. Neurophysiol.*, 4 (1952), pp. 443-456.
59. Lisker, L., "Linguistic segments, acoustic segments, and synthetic speech", *Language*, 33 (1957), pp. 370-374.
60. Livingston, W. K., Haugen, F. P. and Brookhart, J. M., "Functional organization of the central nervous system", *Neurology*, 4 (1954), pp. 485-496.

61. Malécot, A., "Acoustic cues for nasal consonants", *Language*, 32 (1956), pp. 274-284.
62. ——, "Nasal syllabics in American English", *J. Speech Hearing Res.*, 3 (1960), pp. 268-274.
63. Miller, G. A., *Language and Communication* (New York, McGraw-Hill, 1951).
64. ——, "Information theory and the study of speech", in *Current Trends in Information Theory* (Pittsburgh, U. of Pittsburgh Press, 1953).
65. ——, Bruner, J. S. and Postman, L., "Familiarity of letter sequences and tachistoscopic identification", *J. gen. Psychol.*, 50 (1954), pp. 129-139.
66. ——, and Licklider, J. C. R., "The intelligibility of interrupted speech", *J. acoust. Soc. Amer.*, 22 (1950), pp. 167-173.
67. ——, and Taylor, W. G., "The perception of repeated bursts of noise", *J. acoust. Soc. Amer.*, 20 (1948), pp. 171-182.
68. Monsees, Edna K., "Aphasia in children, diagnosis and education", *Volta Rev.*, 59 (1957), pp. 392-401, 414.
69. ——, "Aphasia in children", *J. Speech and Hearing Dis.*, 26 (1961), pp. 83-86.
70. Morley, Muriel E., *The Development and Disorders of Speech in Children* (Edinburgh, E. & S. Livingstone, 1957).
71. Neff, W. D., "Temporal pattern discrimination in lower animals and its relation to language perception in man", in A. deReuck and M. O'Conner (eds.), *Disorders of Language* (Boston, Little, Brown, 1964).
72. Orton, S. T., *Reading, Writing and Speech Problems in Children* (New York, W. W. Norton, 1937).
73. Peterson, G. E., "An oral communication model", *Language*, 31 (1955), pp. 414-427.
74. Piéron, H., *The Sensations* (New Haven, Yale University Press, 1952).
75. Poggio, G. F. and Mountcastle, V. B., "A study of the functional contributions of the lemniscal and spinothalamic systems to somatic sensibility", *Bul. Johns Hopkins Hospital*, 106 (1960), pp. 266-316.
76. Rosenblith, W. A., *Sensory Communication* (New York, John Wiley, 1961).
77. Rosenzweig, M. R. and Rosenblith, W. A., "Some electrophysical correlates of the perception of successive clicks", *J. acoust. Soc. Amer.*, 22 (1950), pp. 878-880.
78. Sandel, T. T. and Stevens, K. N., "Effect of intensity, duration and spectrum on the perception of two-element auditory sequences", *J. acoust. Soc. Amer.*, 30 (1958), p. 674 (Abstract).
79. Saporta, S., "Frequency of consonant clusters", *Language*, 31 (1955), pp. 25-30.
80. Shannon, C. E. and Weaver, W., *The Mathematical Theory of Communication* (Urbana, U. of Ill. Press, 1949).
81. Shepherd, G., "Studies in tachyphemia: II. Phonetic description of cluttered speech", *Logos*, 3 (1960), pp. 73-81.
82. Silversten, Eva, "Segment inventories for speech synthesis", *J. acoust. Soc. Amer.*, 32 (1960), p. 919 (Abstract).
83. Small, A. M. Jr. and Campbell, R. A., "Temporal differential sensitivity for auditory stimuli", *Amer. J. Psychol.*, 75 (1962), pp. 401-410.

84. Sperry, R. W., "Neurology and the mind-brain problem", in R. L. Isaacson (ed.), *Basic Readings in Neuropsychology* (New York, Harper and Row, 1964).
85. Stroud, J. M., "The fine structure of psychological time", in H. Quastler (ed.), *Information Theory in Psychology* (Glencoe, Ill., The Free Press, 1955).
86. Sturtevant, E. H., *Linguistic Change* (Chicago, U. of Chicago Press, 1917).
87. ——, *An Introduction to Linguistic Science* (New Haven, Yale U. Press, 1947).
88. Tiffany, W. R., "The threshold reliability of recording sustained vowels", *J. Speech Hearing Dis.*, 18 (1953), pp. 379-385.
89. Tobias, J. V. and Zerlin, S., "Lateralization threshold as a function of stimulus duration", *J. acoust. Soc. Amer.*, 31 (1959), pp. 1591-1594.
90. Twaddell, W. F., "Phonemes and allophones in speech analysis", *J. acoust. Soc. Amer.*, 24 (1952), pp. 607-611.
91. Van Riper, C. and Irwin, J. V., *Voice and Articulation* (Englewood Cliffs, N. J., Prentice-Hall, 1958).
92. Wallach, H., Newman, E. B., and Rosenzweig, M. R., "The precedence effect in sound localization", *Amer. J. Psychol.*, 62 (1949), pp. 315-336.
93. Warren, R. M. and Gregory, R. L., "An auditory analogue of the visual reversible figure", *Amer. J. Psychol.*, 71 (1958), pp. 612-613.
94. Weiner, N., "Time, communication, and the nervous system", *Ann. N. Y. Acad. Sci.*, 50, Art. 4 (1948), pp. 197-220.
95. ——, "Speech, language and thought", *J. acoust. Soc. Amer.*, 22 (1950), pp. 696-697.
96. West, R., Annsberry, M., and Carr, Anna, *The Rehabilitation of Speech* (New York, Harper, 1957).
97. Wever, E. G., *Theory of Hearing* (New York, John Wiley, 1949).
98. White, C. T. and Cheatham, P. G., "Temporal numerosity: IV. A comparison of the major senses", *J. exp. Psychol.*, 58 (1959), pp. 441-444.
99. Zipf, G. K., *The Psycho-Biology of Language* (Boston, Houghton Mifflin, 1935).
100. Zwislocki, J., "Theory of temporal auditory summation", *J. acoust. Soc. Amer.*, 32 (1960), pp. 1046-1060.

JANUA LINGUARUM

STUDIA MEMORIAE NICOLAI VAN WIJK DEDICATA

Edited by Cornelis H. van Schooneveld

SERIES MINOR

1. ROMAN JAKOBSON and MORRIS HALLE: Fundamentals of Language. 1956. 97 pp. Gld. 6.—
3. EMIL PETROVICI: Kann das Phonemsystem einer Sprache durch fremden Einfluss umgestaltet werden? Zum slavischen Einfluss auf das rumänische Lautsystem. 1957. 44 pp. Gld. 4.—
4. NOAM CHOMSKY: Syntactic Structures. Fifth printing 1965. 118 pp. Gld. 8.—
5. N. VAN WIJK: Die baltischen und slavischen Akzent- und Intonationssysteme. Ein Beitrag zur Erforschung der baltisch-slavischen Verwandtschaftsverhältnisse. 2nd ed. 1958. 160 pp. Gld. 15.—
7. ERNST PULGRAM: Introduction to the Spectrography of Speech. Second printing 1964. 174 pp., 31 figs., 2 tables. Gld. 12.—
8. AERT H. KUIPERS: Phoneme and Morpheme in Kabardian (Eastern Adyghe). 1960. 124 pp. Gld. 16.—
9. A. ROSETTI: Sur la théorie de la syllabe. Deuxième édition, refondue et augmentée. 1963. 43 pp. Gld. 7.50
10. URIEL and BEATRICE WEINREICH: Yiddish Language and Folklore. A Selective Bibliography for Research. 1959. 66 pp. Gld. 6.—
11. E. and K. DELAVENAY: Bibliography of Mechanical Translation. — Bibliographie de la traduction automatique. 1960. 69 pp. Gld. 10.—
12. CARL L. EBELING: Linguistic Units. Second printing 1962. 143 pp. Gld. 12.—
13. SHELOMO MORAG: The Vocalization Systems of Arabic, Hebrew, and Aramaic. Their Phonetic and Phonemic Principles. 1962. 85 pp., 2 folding tables. Gld. 15.—
14. DWIGHT L. BOLINGER: Generality, Gradience, and the All-or-None. 1961. 46 pp. Gld. 5.50
17. W. SIDNEY ALLEN: Sandhi. The Theoretical, Phonetic, and Historical Bases of Word-Junction in Sanskrit. 1962. 114 pp. Gld. 16.—
19. WALERIAN ŚWIECZKOWSKI: Word Order Patterning in Middle English. A quantitative study based on *Piers Plowman* and *Middle English Sermons*. 1962. 114 pp. Gld. 16.—
20. FINNGEIR HIORTH: Zur formalen Charakterisierung des Satzes. 1962. 152 pp. Gld. 15.—

22. E. F. HADEN, M. S. HAN and Y. W. HAN: A Resonance Theory for Linguistics 1962. 51. pp. Gld. 7.—
23. SAMUEL R. LEVIN: Linguistic Structures in Poetry. Second printing 1964. 64 pp. Gld. 8.—
24. ALPHONSE JUILLAND and JAMES MACRIS: The English Verb System. 1962. 81 pp. Gld. 8.—
25. IVAN FONÁGY: Die Metaphern in der Phonetik. Ein Beitrag zur Entwicklungsgeschichte des wissenschaftlichen Denkens. 1963. 132 pp., 5 figs. Gld. 14.—
26. H. MOL: Fundamentals of Phonetics. I: The Organ of Hearing. 1963. 70 pp., 28 figs. Gld. 9.—
27. LÁSZLÓ ANTAL: Questions of Meaning. 1963. 95 pp. Gld. 10.—
29. PUNYA SLOKA RAY: Language Standardization. Studies in Prescriptive Linguistics. 1963. 159 pp. Gld. 16.—
30. PAUL L. GARVIN: On Linguistic Method. Selected Papers. 1964. 158 pp. Gld. 14.—
31. LÁSZLÓ ANTAL: Content, Meaning, and Understanding. 1964. 63 pp. Gld. 8.—
32. GEORGES MOUNIN: La Machine à Traduire. Histoire des problèmes linguistiques. 1964. 209 pp. Gld. 22.—
33. ROBERT E. LONGACRE: Grammar Discovery Procedure. A Field Manual. 1964. 162 pp. Gld. 9.—
34. WILLIAM S. COOPER: Set Theory and Syntactic Description. 1964. 52 pp. Gld. 8.—
35. LUIS J. PRIETO: Principes de noologie. Fondements de la théorie fonctionnelle du signifié. Préface d'André Martinet. 1964. 130 pp. 36 figs. Gld. 18.—
36. SEYMOUR CHATMAN: A Theory of Meter. 1965. 229 pp., many graphs, 2 plates. Gld. 21.—
37. WAYNE TOSH: Syntactic Translation. 1965. 162 pp., 58 figs. Gld. 21.—
38. NOAM CHOMSKY: Current Issues in Linguistic Theory. 1964. 119 pp. Gld. 10.—
39. D. CRYSTAL and R. QUIRK: Systems of Prosodic and Paralinguistic Features in English. 1964. 94 pp., 16 plates. Gld. 12.—
42. MILKA IVIĆ: Trends in Linguistics. Translated by Muriel Heppell. 1965. 260 pp. Gld. 25.—
43. ISTVÁN FODOR: The Rate of Linguistic Change. Limits of the Application of Mathematical Methods in Linguistics. 1965. 85 pp., figs. Gld. 12.—

MOUTON & CO · PUBLISHERS · THE HAGUE